

Privacy Considerations for Data Collection Via Trackable Content on Mobile Storage Devices (Honey Sticks)

AKA the "Found Wallet" Scenario in a Connected World

By Scott Wright

January 15, 2008

Overview

The age-old dilemma of finding a wallet has always been an ethical issue. The finder could at least inspect the wallet without much fear of being discovered, and then decide what to do with it. We like to think most people would try to return it immediately. However, security professionals are often inclined to think the worst of people. It helps when trying to identify risks.

Today, in the digital world, there is a growing risk in keeping (or at least using) "found" devices. But society's awareness of risk lags behind emerging technology, exposing both the individual who loses a device, as well as the Finderⁱ, to new risks they may under-estimate.

With the evolution of "web-friendly" applications and the falling costs of Mobile Storage Devicesⁱⁱ (or Portable Memory Devices), new risks and opportunities for exploiting user behavior are emerging, embedded in everything from research to system attacks.



Mobile Storage Devices

The term ***Honey Stick*** is used in this paper to refer to devices configured and placed in a location for the purpose of enticing an

individual into connecting it to a computer to explore its contents, and surreptitiously gathering information, or launching more malicious activities. The assumption is that the computer used by the Finder will be connected to a network with access to the Internet at some point, to enable communication of information back to the "mother ship".

When a Honey Stick is used for research, or for defense against loss or theft, there is a fine line between the rights of the "Finder" and the rights and obligations of the "Dipper"ⁱⁱⁱ. This paper provides discussion and outlines some safety measures that should be used to minimize the privacy and security risks to both parties.

If the Finder uses the safest possible methods for examining the Honey Stick's contents, there should be no privacy risk, and minimal security risk. However, without explicitly acting to avoid risks, the finder is subjecting themselves to potentially significant risks in the form of viruses, Trojan horse programs, keyloggers and "stick-phishing"^{iv} attacks. Some suggestions are provided below for handling "found" devices in a way that reduces risk, regardless of how they are configured. These methods are not guaranteed to be safe, but are much safer than blindly inserting a device into a computer to examine its contents.

If the Dipper has honest intentions and uses responsible methods, some information can safely be gathered about individuals who find the Honey Stick, without risk of infecting the Finder's computer or adversely impact the finder's privacy.

Some suggested methods are provided in this paper to reduce risks to the Finder, as well as subsequent potential legal or ethical risks to the Dipper for any damages resulting from the Honey Stick's use. The focus here is on privacy issues related to legitimate data collection methods.

Background and Technological Premise

As access to Internet and digital media have become more affordable and popular, it has become easier for anyone with a web site to track their visitors' on-line behavior. From the World-Wide Web's inception, Web servers have been able to obtain a limited amount of data from virtually every request made to it by a browser (with a few exceptions). Each request usually contains, as a minimum, the IP address^v of the requester and the URL of the Web content (or page) being requested.

In addition to this basic information, as a user loads Web pages and clicks on hyperlinks within them, their session's browsing habits can be tracked. It is possible to discover preferences, as well as the collection of user-entered information, via forms. Very often even more tracking information can be gathered using "browser cookies"^{vi}, as well as by leveraging installed plugins that extend the capability of browsers.

The privacy risks related to these mechanisms are well-documented. However, it is getting harder to find sites that work well when cookies and other privacy intrusive features turned off. Some loss of privacy is almost a given with each browser session.

Most often, URLs are requested intentionally by the user, either as published addresses, explicit links from other websites, or explicit links from emails. However, with the development of Web site spoofing, phishing attacks and applications that use embedded HTTP requests, sites are able to discover this information without a user necessarily knowing they have initiated requests to a particular website. Most malicious risks, such as identity theft or password sniffing, have also been fairly well documented, along with recommended safeguards^{vii}.

Linkable Content Availability

There is emerging widespread use of programs such as Email clients, word processors and spreadsheet programs that now handle HTML files and hyperlinks in the content they handle. These "web-friendly" programs provide potential opportunities for researchers and marketers (not to mention malware authors) to extend the reach of tracking to include less traditional and more innovative mechanisms.

Linkable content has also led to the evolution of malicious security threats such as Spyware and Trojan Horse programs that often covertly collect information from user's computer system and transfer it back to a central aggregation or analysis point. While these types of threats have been documented, there is still potential for growth in use of innovative methods for data collection, due to the value of information about personal preferences and characteristics that can be gathered. These security threats are beyond the scope of this paper.

Using Trackable Content in Data Collection

The potentially legitimate practice of statistical data collection relying on the tracking of user inputs for almost any type of web-linked

content may become much more common (i.e. tracking accesses to static files on local file systems, as opposed to websites). An ever-increasing number of “web-linkable” files are authored and distributed in many ways, with both benign and malicious intent.

Web-linked and trackable content provides a range of alternatives for collecting data that can present privacy issues. Motivations for an entity to create content that can be tracked include:

1. **Malicious Intent** - Gathering information in support of planning a malicious and/or illegal attack aimed at either collecting more valuable information, or presenting a Denial of Service (DOS) attack against a target entity. In this case, privacy potentially less serious than the possible outcomes of immediate attacks relying on the same user behavior.
2. **Directed Marketing** - Gathering information in support of legal, yet targeted, business initiatives such as directed marketing campaigns. The debate over whether this represents a legitimate “personalization” initiative or a veiled data collection initiative has not been settled. But if the result gives significant value to users without putting them at risk, then it may be a legitimate activity.
3. **Passive Market Research** - Gathering information in support of research projects or marketing initiatives which aggregate and analyze information and draw conclusions about the impacts of specific variables on behavior. It is even less clear as to whether or not this activity constitutes a privacy risk. It seems unlikely that this activity would stand out as being a significant privacy risk against the current backdrop of other risk scenarios. Nevertheless, certain circumstances (e.g. the firing of an employee who let such information become compromised) may carry legal obligations and liabilities for the data collector.

Approaches to Stealthy Data Collection

Triggers that can enable these information gathering mechanisms include:

1. **Active**, or program-initiated tracking, which depends on having a user cause a program to run on their computer or on a web site server that gathers the information and sends it to an aggregation site.

2. **Malicious Web sites** (via phishing, forged pages or enticing ads) that deceive users into believing they are on a site that they trust, so that they will divulge sensitive or valuable information.
3. **Passive**, or unobtrusive tracking, which only requires commonly pre-installed programs on a computer (typically a Web browser), such that visits and interactions with certain sites can be tracked. Links can exist in these files that cause the application opening them to attempt to make requests to the "mother-ship" or home website.

Items (1) and (2) above can also represent common, potentially dangerous threats that users should always be guarding against. Item (3) is less obvious, and may carry a range of risks from benign to malicious.

Passive tracking tends to be limited by what information a Web server can directly collect, but the mechanism may be more easily deployed, since it only makes use of existing programs, simple data files and network connections. Nevertheless, content packaged the same way can also be used as the launching point for more active or dangerous attacks, given that the media may contain a link to a malicious site, and might behave as a phishing site.

New Risks From Mobile Storage Devices

While the traditional methods of gathering data listed above will likely continue to exist and grow, the growth in small, inexpensive Mobile Storage Devices is opening a new area of security and privacy risks that must be addressed.

Loss or theft of data that exists on a Mobile Storage Devices presents serious security and privacy risks for both individuals and businesses. There are many security technologies entering the market that are aimed at reducing exposure to losses directly related to compromise of sensitive data, and also network intrusion risks caused by the connection of unauthorized devices to corporate networks or home systems. These are generally known as "End-Point Security" solutions.

Privacy Issues from "Found" Mobile Storage Devices

A less obvious set of risks arise when unknown or "found" mobile devices are inserted into a computer. This is the **Honey Stick** scenario. Suddenly, the types of information covered by a Privacy

Policy^{viii} become less clear with questions such as, “If you shouldn’t have accessed it, how can I be held liable?” Nevertheless, this resembles the liability questions around a trespasser injuring themselves on private property. This is one reason why the topic deserves some discussion.

Here are some basic questions that begin to arise as a result of the use of Honey Sticks:

1. Is it a breach of an individual’s privacy if they use a “found” device (i.e. the contents are unknown) and the IP address of the device is tracked by its original owner without their explicit permission?
 - a. If so, how serious?
 - Suggestion - This depends on the location and jurisdiction of privacy laws that apply for a given situation. For example, in the European Union this would likely be considered a privacy violation.
 - b. If so, what action should be taken by the data collector (Honey Stick creator or “Dipper”) to mitigate the risk or potential impact to the user?
 - Suggestion - The Dipper can either structure the process so that the user is first given a description of the data collection process, along with an opportunity to view the privacy policy before explicitly acknowledging and approving of IP Address collection; or the Dipper can refrain from collecting IP Addresses or any other PII under the privacy laws, while still collecting statistical information about clicks until approval is given by the user to collect PII.
2. If an individual clicks on files and links on a “found” device, is the tracking or data matching of the “linked text” a breach of the user’s privacy?
 - a. If so, how serious?
 - Suggestion - Provided that no PII is gathered, this would likely not be considered a privacy violation. However, this depends on the jurisdiction’s privacy laws.

- b. If so, what should be done by the Dipper to mitigate the risk or potential impact to the user?
- Suggestion - If the jurisdiction in which the situation occurs includes this information in its definition of PII, then it should not be included in the data gathered.

Discussion of Privacy Issues

Impacts

Tracking of IP addresses is commonplace for most websites, so the impact of having an IP address known is not in itself a significant privacy issue. Responsible organizations often disclose within their published Privacy Policies that they track IP addresses of visitors “to provide a better level of service” based on traffic sources and usage patterns. Perhaps the most significant impact comes when correlated with records of other activities (also called data matching).

The Organization for Economic Cooperation and Development (OECD) has done a significant amount of work with member countries to develop guidelines for developing Privacy Policies. Guidelines produced by their Directorate of Science, Technology and Industry provide some guidance on how to create Privacy Statements and how information should be treated. In fact, the OECD has published a questionnaire-based Privacy Statement Generator that can be used to assist in creating a Privacy Statement.

These guidelines contain some specific statements about the use of IP addresses. Interestingly, IP addresses are not considered PII if they are only used for technical maintenance of websites and other general purposes. But it states that “...*this does not include personal data used to tailor or modify the content to the specific individual or data used to evaluate, target, profile or contact the individual.*”^{ix}

In the case of a Honey Stick being picked up and used by an employee of an organization on an internal computer system, the use of an IP address being logged is clearly to “evaluate or target” the individual. It may be possible for an employer to identify the user who is violating policies. Given that sanctions may arise from policy violations, an employee may be at risk for dismissal by accessing a Honey Stick if they can be shown to have acted improperly. ***Therefore, even the simplest forms of data collected from the use of Honey Sticks may be enough to significantly impact an individual’s privacy rights.***

Safeguards

Collection in Public

The safest approach for a Dipper is to have a documented policy (such as the one found at <http://www.honeystickproject.com/privacy>) and a set of procedures. This would provide the framework to properly safeguard the IP addresses and other data collected.

Safeguarding PII involves having a Privacy Officer who is responsible and accountable for overseeing the handling of PII. The process should ensure that the information is stored securely with access controls and/or encryption to prevent unauthorized individuals from accessing it. In addition, personnel who do have access to it should be held to a confidentiality standard, and should not be allowed to release, access, modify or erase PII without using an auditable process and having authorization from the designated Privacy Officer. This should reduce the likelihood of damage arising to individuals.

It should be noted, however, that if a Finder uses a device from within a corporate network, the HTTP requests may be logged by a firewall or Web traffic filtering system. So, it is still possible for an employer to identify a specific user whose computer is making requests to specific sites. The only way to prevent employers' firewalls from seeing actual request content such as link text or anchor text is to have the links set up secure sessions using SSL encryption. This requires the use of a server-side SSL certificate which must be purchased and registered through a certificate authority such as Verisign or Entrust.

Internal Collection by Organizations

On the other hand, if a company undertakes this type of data collection internally (e.g. to test staff compliance with policies), it should also have a policy in place to disclose its collection, usage, storage and disposal methods for Personally Identifiable Information (PII)^x belonging to employees, clients, partners, suppliers, etc. Obviously, the policies should be followed and procedures for handling the information should be auditable.

It should be noted that the CEO and CIO of the organization should be made aware of any plans to perform this type of activity with any individuals related to the organization (e.g. staff, customers, etc.)

Hired Dippers and Testers

The situation where privately hired Dippers (in the form of penetration testers) may use Honey Sticks should also be addressed. In this case,

the client organization should require that the tester safeguard the data collected according to its own IT Security and Privacy Policies. Private Dippers should, in turn, require the organization to authorize and indemnify them, provided they are acting under the client's direction, and can demonstrate compliance with the client's policies. Again, the CIO and CEO of the target organization should be made aware of any such plans and engagements.

Scenarios Using Files With Passive Trackable Content

The following scenarios represent different approaches and intents in using **Honey Sticks** for information gathering:

1. **For random targets:** A file, or set of files, is configured with filenames that may appeal to different individuals' interests (with the intent of learning about random individuals' preferences). For example, an HTML file may have a filename of "Cool Stuff.html". Within the file, there could be a link to an image on a website, which triggers a log entry on the webserver's logging system. The following link is an example:

```

```

The above link can be made to have no visible elements within the file being opened, and the link may only be found by looking at the HTML source code. Links can also be added within each file that have text links with labels that have various implications, such as "**Jokes**" or "**Funny Videos**", or even "**Click HERE if you found this device and wish to return it to notify its owner**". As each file is opened, and links are clicked on, HTTP requests are logged by the Dipper's website that can later be analyzed to determine something about the user's intent.

2. **In case of loss or theft:** A file, or set of files that may aid in locating the device. There is at least one service, such as GadgetTrak^{xi} which uses software installed on a Mobile Storage Device, together with software on their website to actively collect and send information back to its owner if it is plugged in by a Finder. However, it may be possible to achieve a similar effect with content that only requires a browser, and does not run a separate program. This may be preferable to using a separate program, since it has virtually no possibility of causing damage

to the Finder's computer when it is invoked, thereby presenting less risk or liability for the Dipper.

3. ***For directed market research (e.g. prospects, private enterprise penetration testing on-site)***): or to learn about the interests or habits of any user that the device might be given to (e.g. trade show give-aways with literature files pre-loaded on the device, testing whether employees use safe computing habits or comply with policies, or testing IT Security awareness training effectiveness). This technique can be taken to an extreme by some marketing and advertising websites that use "Web Bugs"^{xii} as a way to communicate information between multiple sites, often invisibly and without disclosing the nature of information sharing that is occurring under the covers.

Countermeasures Against Passive Tracking

Potential subjects or targets of Honey Sticks (in reality, this could be any individual) should be aware of how to defend against unwanted data collection employing Honey Sticks.

Passive Honey Sticks (with only static file content) not work without Internet connectivity at the time of access. So, simply disconnecting the computer from all networks will prevent the data from being collected as files are accessed.

If removable media drivers are disabled, as in some enterprise user desktop configurations, the data on the Honey Stick will not be accessible.

Nevertheless, this may simply drive employees to take the Honey Stick home to a system that is not locked down. A good security awareness program will advise employees that they should never install devices they find in systems at work or at home.

In fact, in a world with perfect security awareness, Honey Sticks should not really work, because nobody should plug them in! This is why Honey Sticks are a good way to measure security awareness within a community, provided the information collected is treated properly.

With this method, there is virtually no risk of damaging the user's computer, since the only things needed for it to work are static data files, such as HTML or static Web pages.

In the end, if sufficient countermeasures are used, then the data may not be received. It may be impossible to tell whether the user picked up and accessed the Honey Stick through a computer with no external network connection, picked it up and stuffed it in a bag or desk and forgot about it, or threw it in the garbage. With this in mind, the use of Honey Sticks for legitimate purposes provides value only in cases where it doesn't matter if some of the devices are never picked up, used or accounted for. It is still a useful approach for identifying what types of files and links are most often opened when Mobile Storage Devices are found.

Using Active Executable Program Files

For any of the above scenarios it is also possible to use program files that may execute automatically in some cases, or may be launched when a user clicks on them (such as the GadgetTrak approach to finding lost devices). The use of a program allows for collection of much more data, but also increases the risk of adversely impacting the host computer it is running on, and may impact the privacy of the Finder with legal implications for the original owner. This is an area that should be explored in more detail.

Countermeasures Against Active Executable Program-based Tracking

Some anti-virus or anti-spyware programs may detect the programs before they can successfully run. Furthermore, many enterprises now lock down user systems so that they prevent unauthorized programs from running, or prevent removable media from being accessed.

The program will only be able to link or report back to its "mother-ship" if there is Internet connectivity at the time of access, unless it has more complex abilities for monitoring connections and delaying its communications until access exists. In this case, it is more difficult to defend against. Firewalls or other active safeguards may be required to prevent or detect this type of activity.

In the end, disabling removable media drives is the only way to guarantee that such programs (and other more malicious programs) are not able to gain a foothold on these systems.

Questions for Further Discussion

1. Is a Privacy Impact Assessment^{xiii} necessary in all situations where data is collected via Honey Sticks, or just in certain conditions?
 - a. The safest approach is to do a PIA prior to initiating any type of system or process that collects information from individuals to determine the need for privacy safeguards.
2. Where is the line drawn that separates “researching” from “spying” when it comes to tracking user behavior?
 - a. Is a line crossed when a program inserted into an individual’s own Mobile Storage Device sends an HTTP request to its home website when activated to provide information about where the misplaced device is located?
 - What if only a trackable web link is placed in a file that is ultimately processed by a Finder’s browser, as opposed to a program on the Honey Stick?
 - b. Is a line crossed when a program inserted into a Mobile Storage Device that is **intended to be “found”** sends an HTTP request to its home website when activated?
 - c. Is a line crossed when a program inserted into a target individual’s Mobile Storage Device sends an HTTP request to its home website when activated?

In most cases, the answers to these questions depend upon who knows about the research, how it is documented, how the results are used and how PII is collected and disposed of. It may also depend on what actions are performed on the Finder’s computer. It is also not clear how risk and liability are affected when these approaches are used in conjunction with Web Bugs (mentioned earlier) to initiate tracking of user actions from seeded marketing materials. Further exploration of these issues is certainly required to answer them.

Scenarios for Future Discussion

The following paragraphs describe comparative situations or models in which individuals may experience consequences due to the indirect actions of another, whether passive or active. Considering these scenarios may help to establish a gauge for the level of relative risk or impact that may arise from the use of Honey Sticks.

The Lost Wallet

When a finder discovers a wallet, the risk assumed in keeping it, or its contents, is usually assumed to be the risk of being charged with theft. Then, presumably, the finder has to be proven guilty of some illegal act. What laws typically govern the situation, and what are their rights?

What specific scenarios might cause the injury or damage to either party? And how much damage could they cause?

Warning Labels and Website References

In some cases, adding visible warning labels (e.g. "Use at your own risk") or website references (eg. www.honeystickproject.com/privacy) to the external case of Mobile Storage Devices may help to address some of the liability and privacy issues by taking measures to notify the Finder of possible risks. This possibility should be explored further with respect to legal implications. However, for the case of Honey Stick research, putting such markings on the devices would be expected to yield a smaller set of measurable results. Depending on the resources and initial results, this may be the subject of future Honey Stick research.

Conclusions

The most significant risks to Finders are still the potential for malicious software infecting their computers, or launching more elaborate identity theft attacks. However, the same safeguards that mitigate these risks to a Finder will usually also thwart any legitimate Honey Stick data collection techniques.

Finders should not insert any Mobile Storage Devices into systems that have sensitive information on them, and should not open files on any found devices (or links within them). Isolating the computer system from the Internet may reduce the risk, but may still have risks associated with malicious code installation. Ideally, Honey Sticks should not be very useful in a world where safeguards are always used. In time, it may also be possible for Anti-Virus software to detect Honey Stick trackable content, in ways similar to those used by the Web Bug software of Bugnosis.com.

For Dippers, the most significant risks of using Honey Sticks are those of the Finder suffering a loss of privacy that may impact their employment or possibly public embarrassment. In either case, the use

of responsible privacy protection practices akin to those used by any organization to protect its clients' or employees' privacy should provide the necessary safeguards to prevent injury or damage. This implies publishing a privacy policy that discloses how information is collected, stored, used and disposed of.

Based on the findings and review comments resulting from a number of drafts of this paper, the Honey Stick Project plans to undertake an initial research project using passive Honey Stick content. It will collect only user actions and disregarding IP addresses that could be considered PII for the purposes of the project. Aggregated statistical results will be posted on the Honey Stick Project website (www.honeystickproject.com). The intent is to educate individuals, as well as enterprise managers and security professionals, regarding the level of security awareness in the general public.

Disclaimers and Disclosures

Scott Wright, B.A.S.C., M.B.A., is a security management consultant and coach, and is author of the Security Views website (www.securityviews.com).



Mr. Wright offers consulting, coaching and training services in the areas of security management, awareness, policies, metrics, threat and risk assessments, security testing, certification and accreditation, development lifecycle security and other process management security issues.

This paper is not a “guide” for ensuring security of any kind. It is a vehicle for informational discussion. The ideas may be used by readers in enhancing the security of systems that they are ultimately responsible for.

Further information about the Honey Stick Project is available at:
<http://www.honeystickproject.com>

Copyright 2008. Scott Wright. All rights reserved. Please contact the author by email at inquiries@securityviews.com or by phone at 613-693-0997 for permission to reprint, or if you have comments on this paper.

ⁱ “Finder”, for the purposes of this paper, is the name given to an individual who finds something left intentionally or unintentionally by another person.

ⁱⁱ “Mobile Storage Devices” is a term used to describe a class of devices that contain digital memory that can be easily accessed when connected to a computer, such as via a USB connector, or a memory card reader. This includes items such as USB memory sticks, Thumb Drives or Jump Drives, digital cameras, digital picture frames, MP3 players, cell phones, PDAs and even some children’s toys.

ⁱⁱⁱ “Dipper” is a term used in this paper to describe an individual who intentionally leaves a Mobile Storage Device in a location hoping that another individual will find it and connect it to their computer to discover what is in its memory.

**Privacy Considerations for Data Collection Via Trackable Content on Mobile Data Devices
(Honey Sticks)**

^{iv} “Stick Phishing” is a term coined in this paper, referring to the use of Honey Sticks to launch attacks that could also be implemented using normal phishing or “spear phishing” attack methods. The Honey Stick is simply a new vector for getting users to a malicious or spoofed Web-site.

^v IP Address – the Internet Protocol address (in the form of 127.0.0.1) that uniquely identifies a network device on the Internet. Note that in some jurisdictions, such as the EU, a user’s IP address is considered Personally Identifiable Information (PII), for the purposes of privacy laws.

^{vi} “Browser Cookies” – small pieces of data deposited by a website to a user’s browser, used to track information that can be linked to that user. For more information please see <http://www.cookiecentral.com> .

^{vii} Phishing risks – Reference on how to defend against phishing and spoofing attacks: <http://www.anti-phishing.info/avoid-phishing.html>, as well as articles on the Security Views website: <http://securityviews.com/blog/category/spam-and-phishing/>

^{viii} “Privacy Policy” – a stated policy published by an organization, stating how that organization collects and treats Personally Identifiable Information. Having a Privacy Policy is a legal requirement in many jurisdictions for organizations or individuals that collect PII.

^{ix} OECD STI Technical Notes on Using the Privacy Statement Generator.

<http://www2.oecd.org/pwv3/help.htm>

^x “Personally Identifiable Information” – any information relating to an individual that could be used to positively identify them.

^{xi} GadgetTrak is a service that uses software loaded on a USB accessible storage device, allowing information about the device to be sent to the owner when it is found and plugged in. The website is located at <http://www.gadgettrak.com/>

^{xii} Web Bugs – a complex arrangement of HTML tags that reference invisible images and browser cookies to transmit information between websites as visitors access their pages. For more information refer to the Bugnosis website at <http://www.bugnosis.org>

^{xiii} Privacy Impact Assessment (PIA) – an analysis tool used to examine the types of information gathered and processed by a system under development or an existing system to determine the extent of risk to Personally Identifiable Information within the it.