Privacy Impact Assessment
for

EINSTEIN 2

May 19, 2008

Contact Point
United States Computer Emergency Readiness Team (US-CERT)
(888) 282-0870

Reviewing Official
Hugo Teufel III
Chief Privacy Officer
Department of Homeland Security
(703) 235-0780

Abstract

This is the Privacy Impact Assessment (PIA) for an updated version of the EINSTEIN System. EINSTEIN is a computer network intrusion detection system (IDS) used to help protect federal executive agency information technology (IT) enterprises. Pursuant to Section 208 of the E-Government Act of 2002 (Public Law 107-347, 44 U.S.C. § 3501, note), the Department of Homeland Security (DHS) must provide this publicly available PIA prior to initiating a new collection of information that uses information technology to collect, maintain or disseminate information that is in an identifiable form or collects identifiable information through the use of information technology. The original PIA for EINSTEIN 1, dated September 2004, explained that EINSTEIN 1 analyzes network flow information from participating federal executive government agencies and provides a high-level perspective from which to observe potential malicious activity in computer network traffic of participating agencies' computer networks.

The updated version, EINSTEIN 2, will incorporate network intrusion detection technology capable of alerting the United States Computer Emergency Readiness Team (US-CERT) to the presence of malicious or potentially harmful computer network activity in federal executive agencies' network traffic. EINSTEIN 2 principally relies on commercially available intrusion detection capabilities to increase the situational awareness of the US-CERT. This network intrusion detection technology uses a set of pre-defined signatures based upon known malicious network traffic. The signatures which will be implemented when EINSTEIN 2 goes "live" are based upon malicious computer code and are not based upon personally identifiable information (PII). Nor is the IDS programmed to specifically collect or locate PII. While future signatures might be developed in response to threats that use what appears to be PII, the purpose of these signatures is to prevent malicious activity from reaching federal networks, not to collect or locate PII. For example, if the author of a computer security exploit chose to use PII in the delivery of malicious code, a signature may be developed in response to that exploit which could contain PII.[1] Accordingly, while the IDS will collect some PII that is directly related to malicious code being transmitted to the federal networks, its main focus is to identify the malicious code and protect federal networks, not to collect PII. In identifying malicious code across the federal networks, EINSTEIN 2 increases situational awareness and provides an improved real-time ability to address computer network incidents on federal systems.

Overview

Protecting the federal executive agencies' IT infrastructure is a substantial undertaking. Under the Federal Information Security Management Act of 2002 (FISMA) (44 U.S.C. § 3541 *et seq.*), all federal departments and agencies must adhere to information security best practices. As such, federal departments and agencies use individual intrusion detection systems to help protect their own computers, networks, and information. Within the National Cyber Security Division of the Department

[1] For example, the Melissa virus (http://www.cert.org/advisories/CA-1999-04.html) propagates in the form of an email message containing malicious code as an attachment. That email message could contain PII.

of Homeland Security, the US-CERT serves as a focal point for addressing computer network security incidents within the federal government. One of the primary functions of the US-CERT is to increase the federal government's awareness of computer network threats and vulnerabilities thereby increasing the government's ability to prepare for and respond to computer network security events.

To improve the US-CERT's capability to maintain situational awareness, all federal executive agencies[2], in accordance with the Office of Management and Budget (OMB) November 20, 2007, Memorandum M-08-05, Implementation of Trusted Internet Connection, will be required to use EINSTEIN 2. This expanded use of EINSTEIN 2 enables the US-CERT to gain increased situational awareness from all the federal executive agencies and fulfill its mandate to act as a central point for computer network security of the federal enterprise.

EINSTEIN 1, developed in 2003, provides an automated process for collecting, correlating, and analyzing computer network security information from voluntary participating federal executive agencies. It works by collecting network flow records. "Flow records" are records of connections made to a federal executive agency's IT systems. The records identify: the source Internet Protocol (IP) address of the computer that connects to the federal system; the port the source uses to communicate; the time the communication occurred; the federal destination IP address; the protocol used to communicate; and, the destination port. Using network flow records, the US-CERT can detect certain types of malicious activity and coordinate with the appropriate federal executive agencies to mitigate those threats and vulnerabilities. The US-CERT shares this analysis, along with additional computer network security information, with both the public and private sectors, via its web site.

EINSTEIN 2, like EINSTEIN 1, will continue to passively observe network traffic to and from participating federal executive agencies' networks. In addition, EINSTEIN 2 will alert when specific malicious network activity is detected and provide the US-CERT with increased insight into the nature of that activity. Through EINSTEIN 2, the US-CERT will be able to analyze malicious activity occurring across the federal IT networks resulting in improved computer network security situational awareness. This increase in situational awareness can then be shared with federal executive agencies in an effort to reduce and prevent computer network vulnerabilities.

EINSTEIN 2 adds to EINSTEIN 1 a network intrusion detection technology that will monitor for malicious activity in network traffic to and from participating federal executive agencies. EINSTEIN 2's network intrusion detection technology uses a set of pre-defined signatures based upon known malicious network traffic.

Signatures are specific patterns of network traffic that affect the integrity, confidentiality, or availability of computer networks, systems, and information. For example, a specific signature might identify a known computer virus that is designed to delete files from a computer without authorization. Signatures are derived from numerous sources such as: commercial or public computer security information; incidents reported to the US-CERT; information from federal partners; or, independent in-depth analysis by the US-CERT. As mentioned above, the signatures which will be implemented when EINSTEIN 2 goes "live" are based upon malicious computer code and are not based upon PII. Nor is the IDS programmed to specifically collect or locate PII. While future signatures might be developed in response to threats that use what appears to be PII, the purpose of these signatures is to prevent malicious activity from reaching federal networks, not to collect or locate PII. For example, if a computer security exploit chose to use PII in the delivery of malicious code, a signature may be developed in response to that exploit which could contain PII.[3] Accordingly, while the IDS will collect some PII that is directly related to malicious code being transmitted to the federal networks, its main focus is to identify the malicious code and protect federal networks, not to collect PII. All signatures will be reviewed by the US-CERT in accordance with legal and privacy guidelines before being employed.

EINSTEIN 2 will alert the US-CERT when the system identifies malicious network traffic occurring in a federal executive agencies' network in response to specific predefined signatures. EINSTEIN 2 sensors only monitor for specific predefined signatures of known malicious activity. EINSTEIN 2 does not seek or obtain the content of all electronic communications. Rather, by scanning communications during transmission, EINSTEIN 2 identifies harmful communications that warrant analysis. A US-CERT analyst may then query that specific information in EINSTEIN 2 to analyze the potentially harmful network traffic identified by the alert. The US-CERT analysts will view only the specific intrusion detection information that caused the triggering alert. The intrusion detection information used by the US-CERT is that portion of the network traffic that is relevant to the specific signature, along with the network traffic that is reasonably related to and associated with the network connection that caused the triggering alert.

EINSTEIN 2 is to augment--not replace or reduce--the current computer network security practices of participating federal executive agencies. Participating agencies will continue to operate their own intrusion detection and prevention systems, perform network monitoring, and use other information security technologies. EINSTEIN 2 enables the US-CERT to correlate activity across the entire federal enterprise. With the enhanced correlation capability, the US-CERT achieves increased situational awareness of federal executive agency computer networks which is required to perform the computer network security responsibilities assigned to DHS.

[2] Not to include Department of Defense or Intelligence Community Executive Branch agencies.

[3] For example, the Melissa virus (http://www.cert.org/advisories/CA-1999-04.html) propagates in the form of an email message containing malicious code as an attachment. That email message could contain PII.

Section 1.0 Characterization of the Information

The following questions are intended to define the scope of the information requested and/or collected as well as reasons for its collection as part of the program, system, rule, or technology being developed.

1.1 What information is collected, used, disseminated, or maintained in the system?

The information collected, used, disseminated or maintained is information derived from communications made to and from the federal networks. This will include communications sent to the federal networks by the public and those communications generated by users of the federal networks. The information collected takes the "form" of network flow records and network packets collected in response to alerts triggered by pre-determined intrusion detection signatures. When malicious traffic triggers an alert, that data will be captured along with the data that is transmitted in proximity to that alert and related to that connection. When data is captured due to an alert being triggered, there is a slight risk that personal information may be transmitted along with a malicious activity. It is the malicious activity that the IDS is focused on and not the PII. EINSTEIN 2 will maintain this captured information on a separate network under the control of the US-CERT. The US-CERT may disseminate this information with federal executive agencies according to written standard operating procedures. The method EINSTEIN 2 uses to collect the information is set forth below.

Client – Server Model & Flow Records

Under the client/server model each entity connected to the Internet is assigned an IP addresses which permits other connected entities to send it communications. This is typically known as the client/server model of information delivery. Typically, the client is a desktop computer or the software that runs on it and the server, also known as the host, is the more powerful computer that houses the data and/or server software. The connection to the server can occur many ways, via LAN, phone line, cable, or modem. In the case of the Internet's World Wide Web, the client is actually the browser on your PC and the server is a host computer located somewhere on the Internet. Typically the browser sends the server a request for a Web page. The server processes that request and sends the answer back to the browser. The connection between the client and the server is maintained only during the actual exchange of information (the connection). Thus after a Web page is transferred from the server, the connection between that computer and the client is broken. Browsers interact with the server using a set of instructions called protocols. These protocols help in the accurate transfer of data through requests from a browser and responses from the server. There are many protocols available on the Internet. The World Wide Web, which is a part of the Internet, brings all these protocols under one roof.[4]

[4] Gralla, Preston, *How the Internet Works*, p. 19 (Que Publishing 2004); *see also, In Re Doubleclick, Inc.*, 154 F. Supp 497 (S.D. NY 2001)(Technology required to communicate with Internet described in detail.)

Flow is a computer network traffic summarization format widely used by network engineers and security analysts. It summarizes communication between two hosts communicating over the Internet. A flow record is created from multiple, related packets grouped together under a common label. This record stores the source and destination IP address; source and destination port; the IP protocol; and associated derived metrics such as timing information and traffic volumes. No packet payload is stored in a flow record. Conceptually, a flow record is akin to a telephone call record -- details such as the caller's phone number and length of the call are stored, but the contents of the conversation are not.

Signatures

As noted in the Overview, EINSTEIN 2 also uses a signature-based detection method. A signature, as defined by in NIST Special Publication 800-94, is a “pattern that corresponds to a known threat.” The NIST Special Publication 800-94 provides that:

> Signature-based detection is the process of comparing signatures against observed events to identify possible incidents. Examples of signatures are as follows:
>
> - A telnet attempt with a username of “root”, which is a violation of an organization’s security policy
> - An e-mail with a subject of “Free pictures!” and an attachment filename of “freepics.exe”, which are characteristics of a known form of malware
> - An operating system log entry with a status code value of 645, which indicates that the host’s auditing has been disabled.
>
> Signature-based detection is very effective at detecting known threats but largely ineffective at detecting previously unknown threats, threats disguised by the use of evasion techniques, and many variants of known threats. For example, if an attacker modified the malware in the previous example to use a filename of “freepics2.exe”, a signature looking for “freepics.exe” would not match it.

The US-CERT uses known patterns of malicious activities to create signatures for inclusion in EINSTEIN 2’s intrusion detection capabilities. The US-CERT will also implement a review process for all new signatures to ensure that the signatures are narrowly tailored to specific computer network activities. This process includes a specific review to ensure that the new signature actually identifies malicious activity and only a minimal amount of raw network traffic is captured to properly identify the computer network event.

Anomaly-based Detection

In addition, EINSTEIN 2 uses anomaly-based detection methods to identify harmful or malicious computer network incidents. Anomaly-based detection, as defined in NIST Special Publication 800-94, is defined as "the process of comparing definitions of what activity is considered normal against observed events to identify significant deviations."

Anomaly detection can best be viewed as an alarm for strange system behavior.[5] It is an activity profile of normal usage over an interval of time. Anything that deviates from the baseline, or the norm, is logged as anomalous. Anomaly detection can be based upon statistical, characteristic, behavioral, protocol or traffic information. Again, the fundamental component to anomaly detection technique is the baseline, or profile. It requires knowing what the normal characteristics of the system are. While an IDS uses a defined set of rules or filters that have been crafted to catch a specific, malicious event, the EINSTEIN 2 anomaly detection capability utilizes the network flow data and alerts to focus on the system's baseline of normal activity. As described above, behavior that varies from this standard is noted. Intrusion detection systems look for a misuse signature and anomaly detection looks for a strange event.[6]

1.2 What are the sources of the information in the system?

As mentioned in Section 1.1, the source of the information collected is that of the network connections established under mechanisms such as the client-server method. The following is an example of a flow record:

Sample Flow record:

```
127.0.0.1|192.168.0.20|52119|25|6|10|600|S|2008/04/28T00:02:47.958|44.9
85|2008/04/28T00:03:32.943|SENSOR1|out| S|
sIP|dIP|sPort|dPort|protocol|packets|bytes|flags|sTime|dur|eTime|sensor
|type|initialFlags|

Explanation of Sample Flow Record:
127.0.0.1 (sIP) IP of Computer who is the source of the connection

192.168.0.20 (dIP) IP of the computer who is the destination of the
connection

52119 (sPort) Port the connection was initiated on by the source
computer

25 (dPort) Port the connection was received on by the destination
computer

6 (protocol) Protocol number, the number is based on the protocol being
used to transport the data (6 = TCP, 1 = ICMP, 17 = UDP)

10 (packets) Count of total number of packets seen in this single
connection (calculated by the sensor)
```

[5] The concept stems from a paper fundamental to the field of security - An Intrusion Detection Model, by Dorothy Denning, http://www.cs.georgetown.edu/~denning/infosec/ids-model.rtf (last viewed May 7, 2008).

[6] *See* http://www.securityfocus.com/infocus/1600 (last visited May 7, 2008).

600 (bytes) Count of total number of bytes seen in this single connection (calculated by the sensor)
S (flags) Aggregation of all flags seen in this single connection. Flags describe what happened in the connection

2008/04/28T00:02:47.958 (sTime) Start time of the connection, Universal Timestamp added by sensor to indicate when the connection was started

44.985 (dur) Duration of the connection, this field is calculated (dur = eTime - sTime)

2008/04/28T00:03:32.943 (eTime) End time of the connection, Universal Timestamp added by sensor to indicate when the connection was ended

SENSOR1 (sensor) Name of the Sensor that collected the data, this field is added by the sensor

out (type) Direction of the traffic (types include “in,inweb,inicmp,out,outweb,outicmp, int2int,ext2ext")

S (initialFlags) First flag seen in the connection, this is only based on the first packet of the connection

Flag Markers and their meanings
C = CWR - Congestion Window Reduced
E = ECE - Explicit Congestion Notification echo U = URG - Urgent A = ACK - Acknowledgement P = PSH - Push R = RST - Reset S = SYN - Synchronize F = FIN – Finished

Additionally, intrusion detection information will be collected in response to alerts from developed signatures. For illustrative purposes only, the following is an example of a commercially available signature. (This is not a signature the US-CERT intends to use.)

```
alert tcp any any -> $HOME_NET 443 (msg:"DoS Attempt";
flow:to_server,established; content:"|16 03 00|"; offset:0; depth:3;
content:"|01|"; within:1; distance:2; byte_jump:1,37,relative,align;
byte_test:2,>,255,0,relative; reference:cve; classtype:attempted-dos;
sid:2000016; rev:5;)
```

Explanation of Signature:

Alert: Type of IDS Event

tcp: Protocol being examined

any: Any source IP

any: Any source port

->: Direction (points to @HOME_NET which indicates inbound)

$HOME_NET: A variable which is defined by the IDS as the subnets that make up the internal network

```
443:  Destination port traffic is bound for

msg:"DoS Attempt":  Name of the alert that is sent to the console (for
humans reading the alert console)
```

The remaining fields of the string tells the IDS what to look for, the breakdown of the commands and instructs the IDS where in the packet to look for the text.

This signature example tells the IDS to alert on any external IP on any external port that sends traffic to the home network, on port 443, with the text "|16 03 00|", and the text "|01|" within certain parameters and offsets. The alert name is defined as "Dos Attempt" and references CVE, SID:2000016, revision 5.

1.3 Why is the information being collected, used, disseminated, or maintained?

The purpose of EINSTEIN 2 is to provide increased computer network security through detecting malicious activities occurring on federal executive agency computer networks. The US-CERT will use this information to fulfill its responsibilities to analyze and reduce computer network threats and vulnerabilities; disseminate computer network security threat warning information; and, coordinate incident response activities.

1.4 How is the information collected?

The EINSTEIN 2 sensor consists of a computer configured with commercial off the shelf software, government developed software, and commercial intrusion detection software. It will be deployed at participating federal executive agencies' Internet Access Points. It is envisioned that these access points will be those being promoted under the OMB Trusted Internet Connection initiative.[7] At these access points the EINSTEIN 2 sensor obtains the network flow information as indicated above. Additionally, EINSTEIN 2 analyzes computer packets as they are being transmitted to and from the federal agency's networks. If these packets match the patterns of the intrusion detection signatures an alert is triggered in which those packets, and those packets that are reasonably related to the connection, are captured for analysis of the computer network incident.

1.5 How will the information be checked for accuracy?

The hardware and software of this system are not programmed to manipulate or modify any data. EINSTEIN 2 maintains exact copies of intrusion detection information transmitted to or from the

[7] EINSTEIN 2 is associated with the Trusted Internet Connection initiative (see OMB M-08-05), but will not strictly be limited to use at Trusted Internet Connections. The US-CERT will seek to maximize the efficiency and benefit from EINSTEIN 2 by focusing on networks containing aggregated Internet traffic to and from participating federal agencies.

federal network. For example, if a connection "spoofs" an IP address (manipulates the data packets it transmits to the federal network to appear as being sent from one source when they come from another source) the intrusion detection system will simply record those packets with the "spoofed" IP address.

1.6 What specific legal authorities, arrangements, and/or agreements defined the collection of information?

EINSTEIN 2 furthers the Department's network security and critical infrastructure protection responsibilities assigned in the Homeland Security Act, FISMA, and related authorities. *See* 6 U.S.C. §§ 101 *et seq.* and 44 U.S.C. §§ 3541 *et seq.*. Moreover, all federal executive agencies, in accordance with the Office of Management and Budget (OMB) November 20, 2007, Memorandum M-08-05, Implementation of Trusted Internet Connection, will be required to use EINSTEIN 2. As such, the US-CERT will enter into a Memorandum of Understanding with each participating agency articulating the specific services the US-CERT will provide through EINSTEIN 2.

1.7 Privacy Impact Analysis: Given the amount and type of data collected, discuss the privacy risks identified and how they were mitigated.

EINSTEIN 1 collects flow record information, which is limited to a small subset of data fields focused on the technical details of network transactions between computers. Flow record data includes IP addresses but does not contain any additional information to identify the individuals communicating. The flow record data is stored in a government-operated, -owned, or -leased secured facility and is only reviewed by the US-CERT.

As the first government-wide intrusion detection system, EINSTEIN 2 will analyze and obtain more network traffic than the federal government has previously been able to rely on to assess threats to federal networks. In addition, the amount of network flow record data (information described above) being captured will increase as more federal agencies are monitored by EINSTEIN 2. Additionally, EINSTEIN 2, as an intrusion detection system, also observes and analyzes all network traffic that connects to a federal executive agency IT system. When malicious traffic triggers an alert, that data will be captured along with the data that is transmitted in proximity to that alert and related to that connection. When data is captured due to an alert being triggered, there is a slight risk that personal information may be transmitted along with a malicious activity. This risk is initially mitigated by establishing specific rule-based signatures developed to identify specific malicious activity. EINSTEIN 2 will use the minimal amount of signatures necessary to effectively defend the federal executive agencies' IT networks. Secondly, the privacy risk is mitigated by limiting how the intrusion detection information is viewed. Under EINSTEIN 2 the captured data is only accessed by the intrusion detection computer program. The only detailed computer network traffic data that analysts will see will be the limited portions of the traffic that is specifically tailored to support an alert of an instance of known malicious activity as defined by a signature, and in those limited situations, only trained US-CERT analysts will view the traffic data. If network traffic does not meet the specific criteria of a specific signature, that network traffic will not be viewed by the US-CERT.

Section 2.0 Uses of the Information

The following questions are intended to delineate clearly the use of information and the accuracy of the data being used.

2.1 Describe all the uses of information.

The flow-records, signatures, alerts, and portions of network traffic containing identified malicious activity will be used by trained US-CERT analysts to identify and respond to computer network security incidents and anomalies, improve network security, generate reports for distribution to participating agencies and other partners, and increase the resiliency of critical, electronically delivered government services. Only information that is directly related to a security incident may be included in any of these products. The US-CERT is a computer network defense and security organization that is responsible for increasing the security of federal systems, not investigating or obtaining attribution for a particular event. Computer network security is, however, accomplished using multiple disciplines to secure the federal network and part of this support is provided by law enforcement, intelligence, and other agencies. These other agencies will be notified when a computer network event occurs that falls under their responsibility. The US-CERT will notify that entity only that the event has occurred and will provide them with contact information so they can coordinate directly with the affected participating federal agency.

2.2 What types of tools are used to analyze data and what type of data may be produced?

EINSTEIN 2 uses commercial and in-house network security tools to identify instances of known malicious activities that are observable at the intersection of the Internet and the computer networks of participating federal executive agencies. The US-CERT analysts will continue to use flow record analysis tools from commercial and government sources. Many of the tools to be used by EINSTEIN 2 will be the same tools that are currently used in consumer-level computer security software and those used by the individual participating federal executive agencies.

Participating federal executive agencies will receive the tools and training (including privacy training) to analyze only the flow record information collected through EINSTEIN 2. The alert information will be used by the US-CERT to support analysis efforts in identifying malicious code and signatures.

2.3 If the system uses commercial or publicly available data please explain why and how it is used.

EINSTEIN 2 will not use commercial or publicly available data about individuals. EINSTEIN 2 will use signatures of known malicious activities. Signatures are derived from numerous sources such as: commercial or public computer security information; incidents reported to the US-CERT; information from federal partners; or, independent in-depth analysis by the US-CERT. All signatures will be reviewed by the US-CERT in accordance with legal and privacy guidelines before being used. Analysts at the US-CERT may combine the EINSTEIN 2 data with other commercial or publicly available data, including information about Internet routes, bandwidth, and outages to create better situational awareness. The US-CERT does not focus on the identities of specific individuals and any data obtained from data providers will be limited to information relevant to the protection of computer networks. The US-CERT does not have an intelligence or law enforcement mission. It is a consumer of computer network security information and as such analyzes computer network security information that has been properly collected in accordance with applicable laws. The US-CERT fuses this information into computer network security products to provide a greater and much needed situational awareness.

2.4 Privacy Impact Analysis: Describe any types of controls that may be in place to ensure that information is handled in accordance with the above described uses.

EINSTEIN 2 acts like a commercially available intrusion detection system. As such, protection is inherently built into the system to minimize the amount of inadvertently acquired personal information. While the network data that traverses the connection to the federal network is copied and fed through EINSTEIN 2, as stated above, the network flow records are stripped down to minimal non-content information. Additionally, the data captured in response to an alert contains only that connection information relevant to the alert. EINSTEIN 2 does not use each and every signature available for monitoring a network system. It monitors for specific signatures related to those malicious activities targeting federal executive agencies' networks. Those signatures are based not on PII but on the malicious activities themselves.

This temporary copy of raw computer network traffic, used only to identify known malicious activities based on signatures, is never viewed by any DHS personnel unless the traffic contains previously defined malicious activity. The raw computer network traffic not containing a malicious activity (i.e., "clean" traffic) is promptly deleted from the system once the analysis of the malicious activity concludes. No computer network traffic will be disrupted. The only information produced by EINSTEIN 2 are high-level records of computer network traffic (flow records); alerts that announce that a particular malicious activity occurred for a particular participating federal executive agency; and, only in those cases, selected portions of the network traffic as defined by the particular signature. This reported information will only be handled by trained and experienced computer network security professionals subject to oversight and audits. The intrusion detection system, as programmed, includes detailed log records which make a record of each command run on the system. This is one of the key control features of this system, ensuring that any unauthorized access to the EINSTEIN 2 data, although unlikely, will be monitored.

The situational awareness information the US-CERT communicates to other agencies and to DHS through summary reports is stripped of any information, such as personally identifiable information, that is not directly related or relevant to the security incident.

Finally, all US-CERT analysts who access data flow records, alerts, and raw computer network traffic will be subject to oversight and will receive annual training from the DHS Privacy Office regarding privacy in general and specific privacy issues related to the US-CERT's computer network defense responsibilities.

Section 3.0 Retention

The following questions are intended to outline how long information will be retained after the initial collection.

3.1 How long is information retained?

Flow records, alerts, and specific network traffic related to an alert will be maintained for up to three years, although limitations on available storage may limit the volume, and therefore time period covered. If at any point in the analysis, the specific network traffic or alert is deemed unrelated or potentially collected in error, it will be deleted.

In all cases of false alerts - an alert generated by non-malicious network traffic - that information will be immediately deleted from the EINSTEIN 2 system. Furthermore, a record will be kept of the deletion and the related signature will be re-evaluated. After re-evaluation, the signature will either be corrected or removed from the system.

3.2 Has the retention schedule been approved by the component records officer and the National Archives and Records Administration (NARA)?

An approval request is in process. The Department of Homeland Security is currently working with the DHS Senior Records Officer to develop a disposition schedule, which will be sent to NARA for approval.

3.3 Privacy Impact Analysis: Please discuss the risks associated with the length of time data is retained and how those risks are mitigated.

The risk associated with the use of this computer network security intrusion detection system is actually lower than the risk generated by using a commercially available intrusion detection system. EINSTEIN 2 does not use each and every signature available for monitoring a network system. It monitors only for those specific signatures related to malicious activities targeting federal executive agencies' networks.

There is a nominal risk during the time an intrusion detection system makes a copy of network traffic data in order to monitor the transmission. The length of time that this raw network traffic is retained and analyzed is minimal. The information generated from flow records does not contain personal information. The information generated from an alert is analyzed and any personal information is minimized promptly. All captured information resides upon a secured system with complete record logging to ensure an audit trail is created. This use of minimization, a secured system, and auditing mitigates the risks associated with this intrusion detection system.

Section 4.0 Internal Sharing and Disclosure

The following questions are intended to define the scope of sharing within the Department of Homeland Security.

4.1 With which internal organization(s) is the information shared, what information is shared and for what purpose?

As part of its computer network security responsibilities, the US-CERT generates reports on topics including general computer network security trends; specific incidents after minimizing PII; and, anomalous or suspicious activity observed on federal networks. Attribution--actually identifying the specific individual or entity that established the network connection that triggered an alert--is not included in the reports. These reports are made available to DHS organizations, including the National Cyber Security Center, and other federal executive agencies through systems such as the US-CERT Secure Portal for their use in infrastructure protection and other computer network security related responsibilities. Computer network security is, however, accomplished using multiple disciplines to secure the federal network and part of this support is provided by law enforcement, intelligence, and other agencies. These other agencies will be notified when a computer network event occurs that falls under their responsibility. The US-CERT will notify that law enforcement or intelligence entity of the event and provided them with contact information so they can coordinate directly with the affected participating federal agency.

The overarching purpose of the EINSTEIN system and the sharing of information are to increase the shared situational awareness and ensure that important cyber security information is shared in a timely and efficient manner.

4.2 How is the information transmitted or disclosed?

As stated above, the information is shared in the form of reports that minimize any PII. This information is transmitted to other federal executive agencies in the form of electronic message alerts; written reports; and, posts to the US-CERT Secure Portal.

4.3 Privacy Impact Analysis: Considering the extent of internal information sharing, discuss the privacy risks associated with the sharing and how they were mitigated.

Only PII that is directly related to a security incident is collected in EINSTEIN 2. As mentioned above, when an alert is triggered based upon a signature, the connection event (communication between two computers) is captured. For example, if the alert is triggered by malicious code contained in an attachment to an email, that email will be captured. Many times the analysis of this event will only require looking at the attachment and not even reviewing the contents of the email. However, sometimes the malicious payload is hidden and delivered via the content (or body) of the email. In those circumstances, the analyst focuses on analyzing the event for the malicious payload, not on any content nor PII contained in the event. Only the US-CERT can see the full details of any PII in the flow records, alerts, and related network traffic. All sharing within DHS is in the form of reports that are designed to minimize the PII contained in the EINSTEIN 2 system.

Section 5.0 External Sharing and Disclosure

The following questions are intended to define the content, scope, and authority for information sharing external to DHS which includes Federal, state and local government, and the private sector.

5.1 With which external organization(s) is the information shared, what information is shared, and for what purpose?

There are two classes of external organizations that receive information derived from the US-CERT's EINSTEIN 2 system.

1. The US-CERT provides a service that allows each participating department to access its own specific EINSTEIN 2 flow records, but not the flow records of other participating departments. In addition, the US-CERT may share the raw computer network information collected through EINSTEIN 2, pursuant to individual signatures, with the specific agency on whose network the malicious activity was discovered. The purpose for sharing this raw information will be limited to furthering the analysis of the specific identified malicious activity.

2. Federal agencies (including participating agencies) are able to obtain access to systems such as a secured US-CERT website that contains trend and summary information on computer network security. This trend and summary information is based in part on EINSTEIN 2 information but does not contain any PII obtained through EINSTEIN 2 that is not directly related to a security incident.

In all cases, the US-CERT will share information with participating agencies for one purpose--to improve computer network security and protection. This includes sharing with other agencies having computer security responsibilities. Computer network security is accomplished using multiple disciplines to secure the federal network and part of this support is provided by law enforcement, intelligence, and other agencies. These other agencies will be notified when a computer network event occurs that falls under their responsibility. The US-CERT will notify that entity of the event and provide them with contact information so they can coordinate directly with the affected participating federal agency.

Again, the overarching purpose of the EINSTEIN system and the sharing of information are to increase the shared situational awareness and ensure that important cyber security information is shared in a timely and efficient manner.

5.2 Is the sharing of personally identifiable information outside the Department compatible with the original collection? If so, is it covered by an appropriate routine use in a SORN? If so, please describe. If not, please describe under what legal mechanism the program or system is allowed to share the personally identifiable information outside of DHS.

As noted above, the US-CERT will execute a Memorandum of Understanding with each agency that seeks to deploy EINSTEIN 2. Pursuant to the Memorandum of Understanding, the US-CERT will only share raw computer network information containing PII with the related participating agency for purposes of further analyzing specifically identified malicious activity observed on that agency's computer network

5.3 How is the information shared outside the Department and what security measures safeguard its transmission?

See Section 5.1, above. Two-factor authentication (e.g., a password and a physical token) is required for access to the EINSTEIN 2 flow records by participating agencies and for access to portals and other systems that contain more detailed trend and computer network security information. EINSTEIN 2 information itself is not shared outside the Department, except in the form of reports on topics including general computer network security trends, specific incidents after minimizing PII, and anomalous or suspicious activity observed on federal networks.

5.4 Privacy Impact Analysis: Given the external sharing, explain the privacy risks identified and describe how they were mitigated.

The US-CERT alone has full access to the collective EINSTEIN 2 data from participating agencies. Organizations other than the US-CERT receive general reports as described previously. The US-CERT does not release any reports containing PII generated from data obtained under the EINSTEIN 2 system unless it is directly related to a security incident.

Section 6.0 Notice

The following questions are directed at notice to the individual of the scope of information collected, the right to consent to uses of said information, and the right to decline to provide information.

6.1 Was notice provided to the individual prior to collection of information?

Yes. Federal agencies are required to post notices on their websites as well as at other major points of entry that computer security information is being collected and their system monitored. Such notices cover intrusion detection systems like EINSTEIN 2. Furthermore, users of federal computer systems are provided with logon banners and sign user agreements that specifically notify them of the computer network monitoring. This Privacy Impact Assessment also serves as a general notice to individuals that network traffic flowing to or from participating federal executive agencies may be collected for computer security purposes.

Participating agencies using EINSTEIN 2 are required to certify to the US-CERT that they have appropriate notices/banners/measures in place to provide individuals with notice that their interaction with federal networks is subject to monitoring for computer network security purposes.

While notice on the web site is provided, the specific requirements of the Privacy Act of 1974 (5 U.S.C. § 552a), do not apply to EINSTEIN 2. The Privacy Act requires any agency that maintains a "system of records" to publish in the *Federal Register* a notice of the existence and character of the system (SORN). § 552a(e)(4). In order to qualify as a record under the Privacy Act, an item must contain information that actually describes the individual in some way.[8]

The Act defines a "system of records" as a group of any records under the control of any agency from which information is maintained and retrieved by PII. § 552a(a)(5). EINSTEIN 2 primarily collects and maintains information based upon signatures generated from computer security events and alerts, as opposed to information that identifies an individual. As described above, in rare cases EINSTEIN 2 will collect information which could identify a person (e.g., an unspoofed email address within header information or other PII within records incidentally collected as part of a security incident), but this latter information will be maintained (indexed) by the security incident, not by the PII. Moreover, EINSTEIN 2 retrieves information (via signatures, analyses and reports) not by PII but by the security event which triggered the alert. It is not sufficient for purposes of the Privacy Act that an agency has the mere capability to retrieve information indexed under a person's name, but the agency must in fact retrieve records in this way in order for a system of records to exist. Only when there is actual retrieval of records

[8] *Tobey v. NLRB*, 40 F.3d 469, 471-73 (D.C. Cir. 1994).

keyed to individuals does the Privacy Act require a SORN.[9] The EINSTEIN 2 system does not maintain or query its data using incidentally collected PII. A SORN is therefore not required.

6.2 Do individuals have the opportunity and/or right to decline to provide information?

Personally identifiable information may be required to process or respond to queries made by individuals to the federal government, but it is not mandatory that an individual produce this information. In this day and age it is assumed computer users are aware that they are voluntarily providing some information to the government when they communicate with it via the Internet. Electronic mail and Internet users have no expectation of privacy in the to/from addresses of their messages or the IP addresses of the websites they visit because they should know that this information is provided to and used by service providers for the specific purpose of directing the routing of information. Like telephone numbers, which provide instructions to the switching equipment that processed those numbers, electronic mail to/from addresses and IP addresses are not merely passively conveyed through third party equipment, but rather are voluntarily turned over in order to direct the third party's servers.[10] EINSTEIN 2 does not solicit or seek PII from individuals; rather, EINSTEIN 2 monitors the voluntarily-initiated connections made to the federal network. Individuals, understanding the nature of how the Internet works, may then decide if they want to transmit information to or from the federal IT network.

In addition, all individuals (employees and contractors) logging into their participating agency’s IT systems will be presented with an electronic notice, banner, that notifies them that government computer systems are monitored. These users can then decide if they wish to use the system or not, and decide what information they want to transmit over the government system.

6.3 Do individuals have the right to consent to particular uses of the information? If so, how does the individual exercise the right?

Yes, to the extent that an individual can decide whether or not to connect to the federal network. Once users decide to interact with a federal executive agency IT system, they are subject to the computer security efforts of the US-CERT and the EINSTEIN 2 system, in addition to any individual computer security programs the participating agencies might have in place.

[9] *Henke v. United States Department of Commerce*, 83 F.3d 1453 (D.C. Cir. 1996) and *see* Office of Management and Budget Privacy Act Implementation - Guidelines and Responsibilities, 40 Fed Reg 28948, 28952 (Jul. 9, 1975).

[10] *See United States v. Forrester*, 512 F.3d 500 (9th Cir. Jan. 7, 2008).

6.4 Privacy Impact Analysis: Describe how notice is provided to individuals, and how the risks associated with individuals being unaware of the collection are mitigated.

Participating agencies using EINSTEIN 2 are required to certify to the US-CERT that they have appropriate notices/banners/measures in place to provide individuals with notice that their connection to a federal network is subject to monitoring for computer network security purposes. In addition, this PIA explains the details of EINSTEIN 2 and the standards that the US-CERT will use to detect malicious activity on the computer networks of participating agencies. As mentioned above, users are expected to possess a rudimentary understanding of how computers communicate and therefore understand the limits of their privacy rights when they voluntarily choose to transmit those communications. Given this understanding, the EINSTEIN 2 process still mitigates any possible risks by: capturing network flow records to which individuals do not have an expectation of privacy since no PII is in the information; and, minimizing the amount of network traffic it captures in response to an alert. Analysts then minimize the PII from the alert information and do not disseminate any products containing EINSTEIN 2 generated PII that is not directly related to a security incident.

Section 7.0 Access, Redress and Correction

The following questions are directed at an individual's ability to ensure the accuracy of the information collected about them.

7.1 What are the procedures that allow individuals to gain access to their information?

As discussed in section 6.1, information is not based on information that identifies an individual, and when in rare cases EINSTEIN 2 will collect information that could identify a person (e.g., an unspoofed email address within header information or other PII within records incidentally collected as part of a security incident), this information will be maintained and indexed by the security incident, not by the PII. Moreover, EINSTEIN 2 retrieves information via signatures, analyses and reports, not by PII, but by the security event which triggered the alert. As such, there is no information about an individual that can be accessed.

Individuals may request other information about EINSTEIN 2 under Freedom of Information Act (5 U.S.C. § 552) and may do so by contacting the DHS FOIA office directly:

FOIA
The Privacy Office
U.S. Department of Homeland Security
245 Murray Drive SW
STOP-0550
Washington, DC 20528-0550
Toll-free: 866-431-0486
Telephone: 703-235-0790
Facsimile: 703-235-0443
Email: foia@dhs.gov

7.2 What are the procedures for correcting inaccurate or erroneous information?

There are no separate procedures for individual correction of information in EINSTEIN 2 since flow records and alerts are generated from exact copies of computer network traffic. The US-CERT analysts are specifically trained, and analysts' use of the system recorded, to ensure that use of EINSTEIN 2 is focused solely on the malicious activity data and not on the personal content of the communications, or to obtain the personal attribution of the source of the malicious activity.

7.3 How are individuals notified of the procedures for correcting their information?

This PIA serves as notice of the EINSTEIN 2 system and associated processes. The EINSTEIN 2 system does not collect information specifically about individuals, only malicious activity occurring on computer networks and as such there are no procedures for correcting information as it exists in EINSTEIN 2.

7.4 If no formal redress is provided, what alternatives are available to the individual?

The US-CERT refers individuals to the FOIA process, as described in Section 7.1.

7.5 Privacy Impact Analysis: Discuss the privacy risks associated with the redress available to individuals and how those risks are mitigated.

The redress procedures fall under the FOIA process. The risks associated with those are the same as for all other FOIA inquiries.

Section 8.0 Technical Access and Security

The following questions are intended to describe technical safeguards and security measures.

8.1 What procedures are in place to determine which users may access the system and are they documented?

Access to EINSTEIN 2 is strictly limited to trained US-CERT personnel who are governed by the US-CERT standard operating procedures.

8.2 Will Department contractors have access to the system?

Yes, US-CERT contractors will have access to the system and are subject to the same training, auditing, and oversight that governs the federal employees assigned to the US-CERT.

8.3 Describe what privacy training is provided to users either generally or specifically relevant to the program or system?

All DHS employees are required to have general privacy training. In addition, US-CERT analysts and other persons who might come into contact with EINSTEIN 2 information will receive annual training on privacy, legal, and policy issues specifically related to EINSTEIN 2. This training will include how to address privacy during the development of new signatures, how to generate a report that minimizes the privacy impact, and how to report when a signature seems to be collecting more network traffic than is directly required to analyze the malicious activity.

8.4 Has Certification & Accreditation been completed for the system or systems supporting the program?

Yes. All new components added to EINSTEIN 2 will be subject to further certification and accreditation.

8.5 What auditing measures and technical safeguards are in place to prevent misuse of data?

The EINSTEIN 2 system is located on a separate firewalled network used only by trained US-CERT personnel for the purposes of detecting malicious computer network activity. All users of EINSTEIN 2 are subject to oversight and must use two-factor authentication to access EINSTEIN 2 data. All external reports are reviewed for appropriate minimization of personally identifiable data.

The EINSTEIN 2 system is designed to use an automated method to apply signatures to the network traffic of participating agencies. If network traffic does not match a signature, it will not be available to a US-CERT analyst, thus significantly minimizing the ability for the US-CERT to misuse the access to the full network traffic of participating federal agencies.

8.6 Privacy Impact Analysis: Given the sensitivity and scope of the information collected, as well as any information sharing conducted on the system, what privacy risks were identified and how do the security controls mitigate them?

The EINSTEIN 2 system will be used solely by trained US-CERT personnel and will be located on a secured computer network, operated within secured physical locations. The use of known signatures of known malicious activity will minimize the raw computer network data available to US-CERT analysts. Finally, an ongoing assessment process will be implemented that will constantly review the signatures and related computer network traffic in order to continually refine (and limit) the amount of data used by EINSTEIN 2 and enhance the precision of the US-CERT's detection of malicious computer network activity.

Section 9.0 Technology

The following questions are directed at critically analyzing the selection process for any technologies utilized by the system, including system hardware, RFID, biometrics and other technology.

9.1 What structured development process was used to develop the system?

The EINSTEIN 2 system is based on the infrastructure, systems, and lessons-learned from EINSTEIN 1. EINSTEIN 2 is subject to the same development process used in EINSTEIN 1.

9.2 How was data integrity and security analyzed as part of the design of the system?

Data integrity and security have been built into the EINSTEIN program from the very beginning. The US-CERT analysts are required to undergo extensive training and background checks to ensure that they conform to the established policies, procedures, and processes required by the US-CERT. Furthermore, the systems that collect the information in EINSTEIN 2 have undergone certification & accreditation and are monitored around the clock for integrity and security. The US-CERT uses two-factor authentication and robust information security practices to maintain the integrity, confidentiality, and availability of the EINSTEIN 2 system.

9.3 What design choices were made to enhance privacy?

The EINSTEIN 2 system was designed to focus strictly on detecting malicious computer network activity in an effort to enhance the integrity, confidentiality, or availability of federal agency information systems. Through the use of approved signatures, EINSTEIN 2 only collects network traffic (which may incidentally contain PII) if it is closely related to malicious computer network activity. There is no opportunity for US-CERT analysts to scan all network traffic and any new signatures must be reviewed for appropriateness before placed into the EINSTEIN 2 system.

The US-CERT analysts use tools designed to focus their attention on the portions of network traffic related to computer network security, not the substance or meaning of a personal electronic communication. Finally, only trained US-CERT analysts have full access to the EINSTEIN 2 system and any information that is disseminated is done so in a summary form designed to minimize the impact on privacy.

Contact Point

Randal Vickers
Deputy Director, US-CERT
(888) 282-0870

Approval Signature

Original signed and on file with the DHS Privacy Office.

Hugo Teufel III
Chief Privacy Officer
Department of Homeland Security