A forensically-sound proxy method to collect network digital evidence

A. Castiglione, G. Cattaneo, A. De Santis
Dipartimento di Informatica
Università di Salerno
Ponte don Melillo, 84084 Fisciano (SA)

Draft del 23/11/2011

Abstract

The present invention is related to a method for allowing a user to remotely collect information that is visible throughout network services such as web pages, chat, documents, photo and video. A proxy server receives input from a remote user that choose the service, the timeframe and the information to collect over the network. The proxy presents a user interface through which the remote user can view the information acquired and allows him to send instructions to the proxy in the collection process. The proxy makes use of cryptographic primitives for integrity and non-repudiation with respect to content and time of the collected information. A detailed and signed report may be generated, archived and/or sent to the user. In a further embodiment the proxy itself can consists of a distributed and coordinated collection of proxies.

Campo di applicazione ed un esempio

L'invenzione è utile nel campo della Digital Forensics. Una definizione della Digital Forensics è la seguente: "The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations. (Digital Forensics Research Workshop I, 2001). In particolare l'invenzione è relata prevalentemente all'aspetto "collection" nella Network Forensics, ovvero nell'ambito di servizi di rete.

Ad esempio, l'invenzione può essere utilizzata per l'acquisizione di evidenze digitali di contenuti pubblicati (pagine HTML) su Internet. Considerando l'estrema volatilità delle informazioni presenti su Internet (possono essere modificate o rimosse in ogni istante) è indispensabile acquisirne una copia per provare davanti ad un giudice che tali informazioni sono state rese disponibili su Intenet, configurando, ad esempio, reati come ingiuria a mezzo stampa, diffamazione, violazione del copyright, ecc. In generale chi ha prodotto e pubblicato dei contenuti su Internet è sempre in grado di rimuoverli a suo piacimento e di eliminarne eventuali tracce.

Descrizione generale

Il metodo prevede l'uso di un proxy che agisce come "terza parte fidata".

Il proxy colleziona evidenze digitali di un insieme di contenuti (ad es., pagine HTML, foto) resi disponibili mediante un servizio di rete (ad es., tramite una pagina web accessibile ad un indirizzo (Universal Resource Locator) identificato sia dal nome che dall'indirizzo IP numerico) nell'istante (data e ora) in cui viene effettuata la richiesta da parte dell'utente.

La Figura 1 riporta l'architettura di riferimento.

Figura 1 : Architettura di riferimento

L'utente U si collega al proxy P e specifica il network service W dal quale intende acquisire i contenuti pubblicati. Il proxy P effettuerà la connessione al network service W agendo come un *Proxy trasparente* per la successiva acquisizione, inoltrando i dati all'utente U dopo aver collezionato una copia del flusso di dati che sono transitati da e per il network service W. Il proxy P acquisisce i dati provenienti dal network service W ed in tempo reale produce informazioni atte a garantire l'integrità e il non ripudio del contenuto. Il risultato dell'intero processo viene memorizzato in un unico contenitore che raccoglie il flusso dei dati e le informazioni relative alla connettività di rete. Viene infine generato un report conclusivo. All'intero contenitore viene firmato digitalmente e viene apposta una marca temporale per certificare il tempo (data e ora) in cui il processo si è svolto.

Il proxy P agisce come un "proxy trasparente" reindirizzando tutto il traffico proveniente dall'utente U verso la destinazione W scelta dall'utente e contemporaneamente rindirizzando tutte le risposte provenienti da W verso l'utente U. La definizione comunemente accettata e fornita nella RFC 2616 di proxy trasparente è: "*A 'transparent proxy' is a proxy that does not modify the request or response beyond what is required for proxy authentication and identification*"

In questo modo l'utente U può accedere al network service W esattamente come se fosse direttamente connesso ad esso senza l'intermediazione del proxy P.

Il proxy può consistere in un insieme di 2 o più proxy, anche distribuiti geograficamente, e amministrati indipendentemente, che sono tra loro cooperanti. Ognuno dei proxy acquisisce parallelamente l'evidenza digitale sulla rete dello stesso network service nello stesso time frame. L'evidenza digitale finale consisterà dell'unione delle varie acquisizioni, ognuna con garanzie di integrità e di non ripudio dei contenuti calcolate dai singoli proxy. Inoltre ognuno dei proxy provvederà ad apporre una marca temporale per certificare l'ora locale in cui è avvenuta l'operazione. Il vantaggio della replicazione è quello dell'affidabilità e della resistenza ad attacchi informatici. Se un proxy non è disponibile oppure è compromesso, il servizio continuerà ad essere utilizzabile e potrà tollerare una minoranza di proxy corrotti.

Lo schema di interconnessione nel caso di proxy multipli è mostrato nella Figura 2.

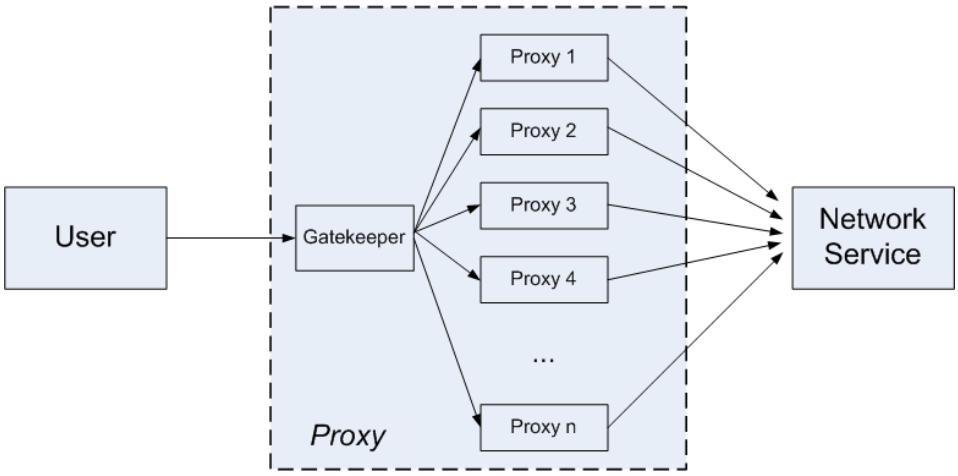

Figura 2: Schema con proxy multipli n >= 2

Un report di tutta l'attività effettuata verrà prodotto dal proxy.

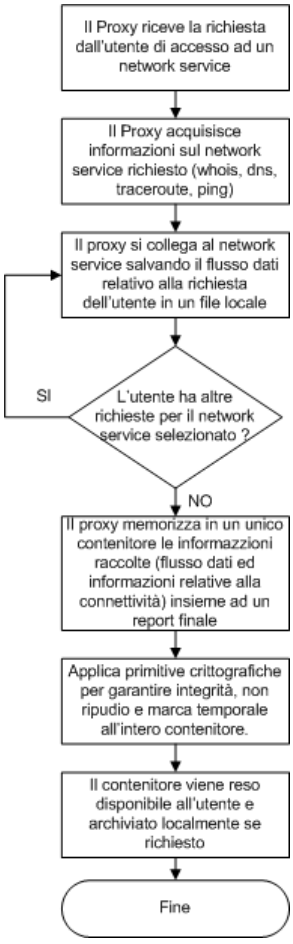

Figura 3: Flow Chart delle attività

Le attività svolte dal proxy sono schematizzate dal flow chart riportato in Figura 3.

Lo schema illustrato nella Figura 4 mostra una possibile applicazione dell'invenzione per l'acquisizione di evidenze digitali relative a servizi erogati attraverso la rete pubblica Internet:

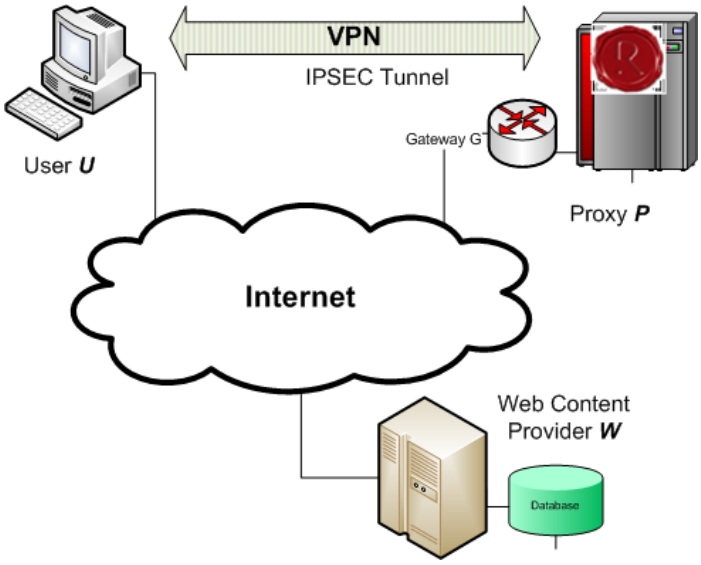

Figura 4: Schema di utilizzo sulla rete Internet

Prior art

Archive.org

- The Internet Archive is a non-profit digital library with the stated mission of "universal access to all knowledge". It offers permanent storage and access to collections of digitized materials, including websites, music, moving images, and nearly 3 million public domain books. The Internet Archive was founded by Brewster Kahle in 1996. (From http://en.wikipedia.org/wiki/Internet_Archive)
- Limitazioni: non per tutti i siti, Facebook, contenuti dinamici. Web 2.0, frequenza di acquisizione, garanzie crittografiche, non è un servizio garantito

Hashbot, by Digital-Security.IT (https://www.hashbot.com/)

- Hashbot is a forensic web tool to acquire and validate, over time, the status of an individual web page or web document. The user inserts the URL to acquire, selects his favorite user agent (default is Firefox) and click on submit. He waits for creating process finish and downloads the zip archive.
- Limitazioni: acquisisce solo pagine html e relativi oggetti, non per tutti i siti, dipende dal browser scelto per l'acquisizione, non supporta contenuti dinamici, non per Facebook, fa solo hash del contenuto, senza firme e timestamp

Packet Analyzers, ad es. Wireshark, tcpdump

- A packet analyzer (also known as a network analyzer, protocol analyzer, or sniffer, or for particular types of networks, an Ethernet sniffer or wireless sniffer) is a computer program or a piece of computer hardware that can intercept and log traffic passing over a digital network or part of a network. As data streams flow across the network, the sniffer captures each packet and, if needed, decodes the packet's raw data, showing the values of various fields in the packet, and analyzes its content according to the appropriate RFC or other specifications. (From http://en.wikipedia.org/wiki/Packet_analyzer)
- Limitazioni se eseguito in locale: installazione del sw, trust di chi effettua l'acquisizione (filtri?), trust del dispositivo che effettua l'acquisizione (troian), garanzie crittografiche
- Limitazioni se eseguito in remoto: trust di chi effettua l'acquisizione (filtri?), garanzie crittografiche

WebCase, by Vere Software (http://www.veresoftware.com/)

- It record any application that appears on the user's computer screen. Access to the program is restricted through the use of a dongle.
- Limitazioni: acquisisce solo ciò che si vede a schermo e ne fa un video oppure cattura lo schermo, non per tutti i sistemi operativi, trust del dispositivo che effettua l'acquisizione (troian), trust di chi effettua l'acquisizione, problemi installamento sw, provenienza oggetti (foto) della pagina, …

Innovazioni rispetto alla prior art

- Trust del proxy che è una terza parte. Soprattutto se è distribuito.
- Non dipende dal sistema operativo e dallo strumento utilizzato per fruire del servizio di rete (ad es., browser) da parte dell'utente. Supporta anche contenuti dinamici.
- Non richiede istallazione di sw sulla macchina dell'utente.
- Acquisisce tutto il flusso di dati a partire dal livello trasporto ai successivi nella stratificazione ISO/OSI. Quindi acquisisce anche quello che l'utente non vede. Per esempio, oggetti della pagina provenienti da sorgenti diverse da quella della URL visibile all'utente.