



中华人民共和国公共安全行业标准

GA/T 1478—2018

法庭科学网站数据获取技术规范

Technical specifications for website data acquisition in forensics

2018-04-13 发布

2018-04-13 实施

中华人民共和国公安部 发布

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国刑事技术标准化技术委员会电子物证检验分技术委员会(SAC/TC 179/SC 7)提出并归口。

本标准起草单位:公安部网络侦察技术研发中心、上海弘连网络科技有限公司、盘石软件(上海)有限公司、厦门美亚柏科信息股份有限公司。

本标准主要起草人:刘晓宇、翟晓飞、陆道宏、宋庆飞、赵庸。

法庭科学网站数据获取技术规范

1 范围

本标准规定了对以 HTTP 和 HTTPS 协议方式提供的网站服务进行数据获取的方法和要求,包括网站服务器基本信息获取以及网站数据内容获取。

本标准适用于法庭科学领域电子物证检验中对网站数据的获取检验。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 5271.1—2000 信息技术 词汇 第1部分:基本术语

GA/T 756—2008 数字化设备证据数据发现提取固定方法

GA/T 976—2012 电子数据法庭科学鉴定通用方法

RFC 959 FTP 协议

RFC 2616 超文本传输协议——HTTP/1.1

3 术语和定义

GB/T 5271.1—2000、GA/T 756—2008、GA/T 976—2012、RFC 959 和 RFC 2616 界定的以及下列术语和定义适用于本文件。

3.1

网站服务器 web server

提供网站内容和网站数据服务的软件环境和硬件设备。

3.2

网站 website

根据一定的规则,使用 HTML 等工具制作的用于展示特定内容的相关网页的集合,包括服务器中的静态文件,或者是服务器按照用户请求动态生成的数据集合。

3.3

静态网页 static web page

在一段时间内,对同一 URL 访问时,服务器返回的数据基本保持不变。包括静态网站页面和服务端端的动态网站页面。

3.4

动态网页 dynamic web page

随着时间的变化,在 URL 保持不变的情况下,网页内容会发生很大的变化。仅指对应于 AJAX、HTML5 等具有客户端动态的页面技术形式。

3.5

富媒体 rich media

在网页中常采用插件的形式,具有动画、声音、视频和/或交互性的信息传播方法。

3.6

用户代理 user agent

一个特殊字符串头,使得服务器能够识别客户使用的操作系统及版本、CPU 类型、浏览器及版本、浏览器渲染引擎、浏览器语言、浏览器插件等。

3.7

网页固定 web page snapshot

对指定的网站页面进行固定,保存为 JPG/PNG/MHT 等存储格式的单一文档。

3.8

网站镜像 website mirror

对网站的页面和文件内容进行保存,修改页面中的链接以保持在本地图面之间的跳转后在本地图面生成的镜像。

4 检验步骤

4.1 时间校对

在网站数据获取时,应与国家授时中心等标准时间源进行时间校对,记录准确的开始时间和结束时间。

4.2 环境建立和基本信息获取

获取并记录相关网站服务器的 IP 地址、协议、服务器软件及版本、域名注册等信息。

4.3 网站获取参数和环境参数记录

4.3.1 网站获取参数记录

在获取网站数据时,宜对获取的深度、关键词、用户、用户 IP 地址、版块、URL、时间、文档类型等参数进行记录。

4.3.2 网站获取环境参数记录

在获取网站数据时,宜对访问时间间隔、代理服务器、用户代理、网站登录和验证码等环境参数进行记录。

4.4 网站数据获取

4.4.1 静态网页数据获取

按照页面链接关系对静态网页内容进行获取,根据需保存网页中的文本内容和图片。静态网页获取形式可采用网页固定、网站镜像、数据内容保存等形式。

4.4.2 动态网页数据获取

对采用 AJAX、HTML5 等技术构建的动态网页进行固定获取,根据需保存页面进程中的数据内容。动态网页获取可采用网页固定和数据内容保存等形式。

4.4.3 数据文件、富媒体应用数据获取

对于网站数据文件以及使用富媒体方式提供的网站内容,应直接下载相关源数据文件,在无法下载的情况下采用截屏、录像的方式进行固定。

4.5 数据内容哈希和校验

对保存的网页固定、网站镜像、数据文件应进行哈希计算,确保页面和文件的完整性。固定的图片页面中应包括与本次固定行为相印证的 URL。

4.6 过程日志记录

网站数据获取过程应采取屏幕录像、摄像和文字记录等方式对关键步骤进行记录。

4.7 检验结果保存

将检出数据采用封盘刻录方式刻录在不可擦写的空白光盘上或者保存在专用存储介质中,并计算检出数据的哈希值。

5 检验记录

检验记录应包括开始时间和结束时间、服务器基本信息、网站获取参数、网站页面数据、文件完整性校验、过程录像和日志记录等。

6 检验结果表述

检验结果表述应符合以下规定:

- a) 检验结果分为检出、未检出、不具备检验条件 3 种;
 - b) 检验结果应根据检验要求对检验对象、检验范围、检验所得进行客观、概括、有针对性的描述;
 - c) 结果表述应包含检材编号、检出情况、检出数据文件或保存检出数据介质哈希值、保存检出数据介质编号等必要信息。
-