普通高等教育“十一五”国家级规划教材

高等院校信息安全专业系列教材

教育部高等学校信息安全类专业教学指导委员会
中国计算机学会教育专业委员会 共同指导

顾问委员会主任：沈昌祥 编委会主任：肖国镇
网络侦查与电子物证系列丛书主编：秦玉海

网站构建分析

肖 萍 主编
刘奇志 徐国天 副主编
秦玉海 审

http://www.tup.com.cn

Information Security

根据教育部高等学校信息安全类专业教学指导委员会编制的
《高等学校信息安全专业指导性专业规范》组织编写

清华大学出版社

普通高等教育“十一五”国家级规划教材

高等院校信息安全专业系列教材

网站构建分析

肖　萍　主编

刘奇志　徐国天　副主编

Information Security

http://www.tup.com.cn

清华大学出版社

北京

内 容 简 介

本书以目前主流的 ASP、PHP 及 JSP 三大网站平台的构建方法为主要讲解方向，以实例分析加案例分析为主要脉络，讲解各类网站平台结构特点及相关调查取证重点命令及方向，包括钓鱼网站的构建与分析过程、在 ASP 网站平台模拟发布敏感信息事件、在 PHP 网站模拟挂马攻击实例等，并对每类案件发生后如何在网站平台下找到相关线索进行了详细剖析。

本书可作为高等院校信息安全等相关专业的教材，也适合从事计算机犯罪现场勘查工作及计算机取证工作的人员，负责企业、公司网站信息安全的从业者，以及对网站平台架构分析技术有兴趣的师生和技术人员参考阅读。

图书在版编目(CIP)数据

网站构建分析/肖萍主编；刘奇志，徐国天副主编. --北京：清华大学出版社，2014
高等院校信息安全专业系列教材
ISBN 978-7-302-34857-3

Ⅰ. ①网… Ⅱ. ①肖… ②刘… ③徐… Ⅲ. ①网站－开发－高等学校－教材 Ⅳ. ①TP393.092

中国版本图书馆 CIP 数据核字(2013)第 314960 号

责任编辑：张　民　赵晓宁
封面设计：常雪影
责任校对：焦丽丽
责任印制：沈　露

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62795954

印 装 者：三河市中晟雅豪印务有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　印　　张：13　　字　　数：299 千字
版　　次：2014 年 1 月第 1 版　　印　　次：2014 年 1 月第1次印刷
印　　数：1～2000
定　　价：25.00 元

产品编号：056294-01

高等院校信息安全专业系列教材

编审委员会

出版说明

21世纪是信息时代，信息已成为社会发展的重要战略资源，社会的信息化已成为当今世界发展的潮流和核心，而信息安全在信息社会中将扮演极为重要的角色，它会直接关系到国家安全、企业经营和人们的日常生活。随着信息安全产业的快速发展，全球对信息安全人才的需求量不断增加，但我国目前信息安全人才极度匮乏，远远不能满足金融、商业、公安、军事和政府等部门的需求。要解决供需矛盾，必须加快信息安全人才的培养，以满足社会对信息安全人才的需求。为此，教育部继2001年批准在武汉大学开设信息安全本科专业之后，又批准了多所高等院校设立信息安全本科专业，而且许多高校和科研院所已设立了信息安全方向的具有硕士和博士学位授予权的学科点。

信息安全是计算机、通信、物理、数学等领域的交叉学科，对于这一新兴学科的培养模式和课程设置，各高校普遍缺乏经验，因此中国计算机学会教育专业委员会和清华大学出版社联合主办了“信息安全专业教育教学研讨会”等一系列研讨活动，并成立了“高等院校信息安全专业系列教材”编审委员会，由我国信息安全领域著名专家肖国镇教授担任编委会主任，共同指导“高等院校信息安全专业系列教材”的编写工作。编委会本着研究先行的指导原则，认真研讨国内外高等院校信息安全专业的教学体系和课程设置，进行了大量前瞻性的研究工作，而且这种研究工作将随着我国信息安全专业的发展不断深入。经过编委会全体委员及相关专家的推荐和审定，确定了本丛书首批教材的作者，这些作者绝大多数都是既在本专业领域有深厚的学术造诣、又在教学第一线有丰富的教学经验的学者、专家。

本系列教材是我国第一套专门针对信息安全专业的教材，其特点是：

① 体系完整、结构合理、内容先进。

② 适应面广：能够满足信息安全、计算机、通信工程等相关专业对信息安全领域课程的教材要求。

③ 立体配套：除主教材外，还配有多媒体电子教案、习题与实验指导等。

④ 版本更新及时，紧跟科学技术的新发展。

为了保证出版质量，我们坚持宁缺毋滥的原则，成熟一本，出版一本，并保持不断更新，力求将我国信息安全领域教育、科研的最新成果和成熟经验反映到教材中来。在全力做好本版教材，满足学生用书的基础上，还经由专家的推荐和审定，遴选了一批国外信息安全领域优秀的教材加入到本系列教

材中，以进一步满足大家对外版书的需求。热切期望广大教师和科研工作者加入我们的队伍，同时也欢迎广大读者对本系列教材提出宝贵意见，以便我们对本系列教材的组织、编写与出版工作不断改进，为我国信息安全专业的教材建设与人才培养做出更大的贡献。

“高等院校信息安全专业系列教材”已于2006年年初正式列入普通高等教育“十一五”国家级教材规划（见教高[2006]9号文件《教育部关于印发普通高等教育“十一五”国家级教材规划选题的通知》）。我们会严把出版环节，保证规划教材的编校和印刷质量，按时完成出版任务。

2007年6月，教育部高等学校信息安全类专业教学指导委员会成立大会暨第一次会议在北京胜利召开。本次会议由教育部高等学校信息安全类专业教学指导委员会主任单位北京工业大学和北京电子科技学院主办，清华大学出版社协办。教育部高等学校信息安全类专业教学指导委员会的成立对我国信息安全专业的发展将起到重要的指导和推动作用。“高等院校信息安全专业系列教材”将在教育部高等学校信息安全类专业教学指导委员会的组织和指导下，进一步体现科学性、系统性和新颖性，及时反映教学改革和课程建设的新成果，并随着我国信息安全学科的发展不断修订和完善。

我们的E-mail地址：zhangm@tup.tsinghua.edu.cn；联系人：张民。

清华大学出版社

前言

随着计算机应用的不断普及，有关计算机犯罪的案例也在不断增多，特别是通过各类网站进行非法活动，如黄色网站、赌博网站、钓鱼网站、在网站上发布虚假敏感信息以及针对网站漏洞进行攻击的案例屡见不鲜。如何能使这类计算机犯罪受到有效的惩治，对计算机犯罪侦查和取证人员提出了更高的要求，其中网站构建基本理论及常见的网站构建技术是必须掌握的内容，只有在熟悉常用的网站构建方法、后台架构、各类网站平台特征、关键的数据库连接文件存放位置、网站数据访问痕迹等问题的基础上才能获得有效的电子证据和案件线索，为打击犯罪提供有利的保障。

本书从计算机犯罪侦查和取证的角度出发，介绍了目前主流的 ASP、PHP 及 JSP 三大网站平台的构建方法、分析了相应特征，并以实例方式介绍了如何追踪保存在网站后台服务器中的数据访问痕迹。

全书共分 6 章。第 1 章是网站构建分析概述，包括网站发展史、工作原理、组成结构、后台架构及实验所用虚拟机软件介绍等。第 2 章是 ASP 网站构建分析，包括典型 ASP 网站构建方法、IDC 网站系统构建及 ASP 网站平台分析。第 3 章是钓鱼网站构建分析，包括网页文件概述、钓鱼网站概述、工作流程、组成结构、构建实例、实例分析等。第 4 章 SQL 注入攻击痕迹分析，包括 SQL 注入成因分析、手工注入实例、读取网站主目录位置、利用差异备份上传网页木马及通过 ASP 注入防火墙来防御 SQL 注入等。第 5 章是 LAMP 平台下 PHP 网站构建分析，主要包括 LAMP 平台简介、PHP 概述、LAMP 平台搭建、LAMP 平台下发布 PHP 网站、LAMP 平台中 PHP 网站分析及 PHP 网站挂马攻击痕迹分析等。第 6 章是 JSP 网站构建分析，包括 JSP 概述、构建 JSP 网站运行平台、发布 JSP 网站及网站分析。

本教材由肖萍负责整体结构设计并编写第 1～第 3、第 5 和第 6 章，徐国天编写了第 4 章，刘奇志编写了第 1.1 节，段严兵编写了第 5.1 节，武晓飞编写了第 3.2 节。本教材的突出特点是实用性强，内容全面，基本涵盖了目前主流的网站开发平台类型。注重理论与实践结合，突出专业特点。

尽管在编写过程中作者做了很多努力，但由于水平有限，教材中不妥之处在所难免，敬请读者批评指正。

编　者

2013 年 12 月

目录

第1章 网站构建概述

1.1 网站发展史

WWW(World Wide Web)是1989年英国人TimBerners. Lee在欧洲共同体的一个大型科研机构任职时发明的。通过Web,互联网上的资源可以在一个网页里比较直观地表示出来;而且资源之间在网页上可以链来链去。在2003年以前的互联网Web应用根据其特点人们通常将其称为Web1.0时代,Web1.0是以静态、单向阅读为主;Web2.0是以分享为特征的实时网络;Web3.0是以网络化和个性化为特征,提供更多人工智能服务。目前的主流模式是Web2.0。

1.1.1 Web1.0

Web1.0时代最明显特征就是用户通过浏览器获取信息,是以网站对用户为主,具体表现如下:

(1) Web1.0基本采用的是技术创新主导模式,信息技术的变革和使用对于网站的新生与发展起到了关键性的作用。新浪公司的最初就是以技术平台起家,搜狐公司以搜索技术起家,腾讯公司以即时通信技术起家,盛大公司以网络游戏起家,在这些网站的创始阶段,技术性的痕迹相当重。

(2) Web1.0的盈利都基于一个共通点,即巨大的点击流量。无论是早期融资还是后期获利,依托的都是为数众多的用户和点击率,以点击率为基础上市或开展增值服务,受众的基础决定了盈利的水平和速度,充分地体现了互联网的眼球经济色彩。

(3) Web1.0的发展出现了向综合门户合流现象,早期的新浪、搜狐、网易等公司,继续坚持了门户网站的道路,而腾讯、MSN、Google等网络"新贵",都纷纷走向了门户网络,尤其是对于新闻信息,有着极大的兴趣。这一情况的出现,使得门户网站本身的盈利空间更加广阔,盈利方式更加多元化,占据网站平台,可以更加有效地实现增值意图,并延伸由主营业务之外的各类服务。

(4) 在Web1.0时代,并不是以html为主,一些动态网站已经被广泛应用,如动网论坛等。

1.1.2 Web2.0

Web2.0是相对Web1.0而言的,和Web1.0相同。它不是一种技术的代名词,而是一个时代的总称。

1. 人是灵魂

在互联网的新时代，信息是由每个人贡献出来的。所有的人共同组成互联网信息源。Web2.0 的灵魂是人。

2. 多人参与

Web1.0 里，互联网内容是由少数编辑人员（或站长）定制的，如搜狐；而在 Web2.0 里，每个人都是内容的供稿者。Web2.0 的内容更多元化：标签 tag、多媒体、在线协作等。在 Web2.0 信息获取渠道里，RSS 订阅扮演者一个很重要的作用。

3. Web2.0 的元素

Web2.0 包含了经常使用到的服务，如博客、播客、维基、P2P 下载、社区、分享服务等。

4. 可读可写互联网

在 Web1.0 里，互联网是“阅读式互联网”，而 Web2.0 是“可写可读互联网”。虽然每个人都参与信息供稿，但在大范围里看，贡献大部分内容的是小部分的人。Web2.0 实际上是对 Web1.0 的信息源进行扩展，使其多样化和个性化。博客是 Web2.0 里十分重要的元素，因为它打破了门户网站的信息垄断，未来博客的地位将更为重要。

1.1.3 Web3.0

Web3.0 的最大价值不是提供信息，而是提供基于不同需求的过滤器，每一种过滤器都是基于一个市场需求。如果说 Web2.0 解决了个性解放的问题，那么 Web3.0 就是解决信息社会机制的问题，也就是最优化信息聚会的问题。

人们只需要输入自己的需求，就可以迅速得到所需信息，甚至是一套完整的解决方案。例如，在计算机中输入：“我想带我 11 岁的孩子去一个温暖的地方度假，我的预算为 3000 美元。”计算机能自动给出一套完整方案，这一方案可能包括度假路线图、适合选择的航班、价格适宜的酒店等。可以预见，承接 Web2.0 的以人为本理念，Web3.0 模式中将会出现各种高度细分领域的平民专家。

真正的 Web3.0 不仅止于根据用户需求提供综合化服务，创建综合化服务平台，关键在于提供基于用户偏好的个性化聚合服务。在 Web3.0 时代，同一模式化的综合服务门户将不复存在，如人们看到的新浪首页将是个人感兴趣的新闻，而那些他不感兴趣的新闻将不会显示。当然，这种个性化的聚合必须依赖强大的智能化识别系统，以及长期对于一个用户互联网行为规律的分析和锁定，它将颠覆传统的综合门户，使得 Web3.0 时代的互联网评价标准不再是浏览的点击率，而是到达率和用户价值。

因此，在 Web3.0 时代能够赢得用户青睐的网站，一定是基于用户行为、习惯和信息的聚合而构建的，人性化、友好界面、简单易用一定是其核心元素，基于用户需求的信息聚合才是互联网的未来发展趋势。

1.2　网站工作原理

在进行网站构建分析之前，需要了解网站的工作模式及访问网站所使用的协议及工作原理。

1.2.1　网站工作模式

目前，大多数网站的工作模式是浏览器和服务器(Browser/Server，B/S)模式。它是对 C/S 结构的一种变化或改进的结构。所谓 C/S(Client/Server)模式，是传统的网络应用程序通常采用的工作模式。这种模式下，客户端机器必须安装特定的客户端应用程序，并做大量的配置工作，才能与指定的服务器进行通信。在 C/S 模式下，系统维护繁琐，维护费用高，而且不易扩展。

在 B/S 模式下，用户工作界面通过浏览器来实现，用户访问网站只需打开 Web 浏览器即可，浏览器类型没有限制，360、TT、IE、Firefox 均可。由于客户端采用的是简单易用的 Web 浏览器软件，不但可以为所有用户提供统一的交互界面，而且也无须像 C/S 模式那样在客户机上安装庞大的客户端应用程序。如图 1-1 所示，B/S 模式通常由 Web 浏览器、Web 服务器和数据库服务器三大部分组成。其中，客户端由 Web 浏览器来实现，它将用户在页面上提交的请求发送给 Web 应用服务器，并将 Web 服务器返回的结果显示给用户。Web 服务器负责接受客户端发过来的页面请求，并将处理结果送回浏览器。数据库服务器的主要任务是根据 Web 服务器发送的请求进行数据库操作(查询、添加、删除与更新等)，并将操作的结果传送给 Web 服务器。

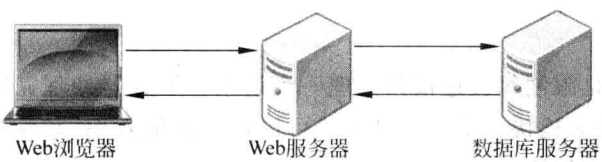

图 1-1　网站工作模式

1.2.2　访问网站所使用的协议

1. HTTP 协议

HTTP 是 Hypertext Transfer Protocol 的缩写形式，称为超文本传输协议，是目前互联网上应用最为广泛的一种网络协议。浏览器请求网页大多数使用 HTTP 协议，浏览器通过 HTTP 协议将网页文件从网站服务器处提取出来的，并翻译成精美漂亮的页面进行显示处理，所有的网页文件都必须遵守这个标准。设计 HTTP 协议的最初目的就是为了提供一种发布和接收 HTML 页面的方法。该协议具有以下特点：

1) 支持请求/响应模式

首先客户端发送一个请求(request)给服务器，服务器在接收到这个请求后将生成一

个响应(response)返回给客户端。这里客户端指终端用户,如 Web 浏览器、网络爬虫或者其他的工具。服务器端是网站,存储着一些资源,如 HTML 文件和图像。客户端与服务器端之间的一次信息交换的完整过程为,首先,客户端与服务器端建立 TCP 连接;然后,客户端发出 HTTP 请求,服务器端发出相应的 HTTP 响应;最后,客户端与服务器端之间的 TCP 连接关闭。

例如,客户想浏览 www.ccpc.edu.cn 网站,则首先浏览器要与网络上域名为 www.ccpc.edu.cn 的 Web 服务器建立 TCP 连接。浏览器发出要求访问 java/book.htm 的 HTTP 的请求。Web 服务器在接收到 HTTP 请求后,解析 HTTP 请求,然后发回包含 book.htm 文件数据的 HTTP 响应。浏览器在接收到 HTTP 响应后,解析 HTTP 响应,并在窗口中展示 book.htm 文件。最后,浏览器与 Web 服务器之间的 TCP 连接关闭。

HTTP 服务器与 HTTP 客户程序分别由不同的软件开发商提供,HTTP 客户程序包括 IE、Netscape 等浏览器;最常用的 HTTP 服务器包括 IIS、Apache 等。HTTP 服务器与 HTTP 客户程序分别由不同的语言编写,并且运行在不同的平台上,双方要看得懂对方发送的数据要归功于 HTTP 协议。HTTP 协议规定了 HTTP 请求和 HTTP 响应的数据格式,HTTP 服务器与客户程序间交换数据都必须遵守 HTTP 协议。

2) 简单快速

客户向服务器请求服务时,只需传送请求方法和路径。请求方法常用的有 GET、HEAD、POST。每种方法规定了客户与服务器联系的不同类型。由于 HTTP 协议简单,使得 HTTP 服务器的程序规模小,因而通信速度很快。

3) 灵活

HTTP 允许传输任意类型的数据对象。正在传输的类型由 Content-Type 加以标记。

4) 无连接

无连接的含义是限制每次连接只处理一个请求。服务器处理完客户的请求,并收到客户的应答后,即断开连接。采用这种方式可以提高传输效率,在没有请求提出时,服务器就不会在那里空闲等待。

5) 无状态

其优点是由于无须记忆状态使得 HTTP 累赘少,系统运行效率高,服务器应答快。

其缺点是由于没有状态,协议对事务处理没有记忆能力,若后续事务处理需要有关前面处理的信息,那么这些信息必须在协议外面保存,导致每次连接需要传送较多的信息。

2. 统一资源定位符(Uniform Resource Locator,URL)

在 HTTP 协议请求数据包中,一个很重要的信息就是 URL,用来指明请求信息的路径,URL 的使用形式如"http://www.ccpc.edu.cn/content.jsp? newsid=9221",通常一个完整的 URL 包括以下 5 部分:

1) 协议

表示客户端与服务端通信采用的协议,通常是 HTTP 协议,但是在一些电子商务网站的关键信息输入网页采用的是 HTTPS 协议,即采用安全技术的 HTTP 协议。例如,

浏览器访问淘宝登录页面，采用的就是 HTTPS 协议。

2）主机

所访问网站的域名或 IP 地址，如果为域名，在真正访问网站服务器前，必须将该域名解析为对应的 IP 地址，域名便于人们记忆，但最终用来寻址定位网站的是 IP 地址。

3）端口

如果为 Web 默认端口 80，则可以在 URL 里面省略，如果网站使用的是非默认端口，则必须在 URL 里指明，如 http://210.47.128.134:8080/index.jsp。

4）文件

客户端要访问的具体网页名称，通常带有路径信息，如果该部分省略，表示访问的为默认文档，默认文档通常为 index.htm、index.asp、index.jsp、index.php 等，具体见 Web 服务器设置。

5）附加资源

URL 字符串中“?”后面的字符串为附加资源，“&”可连接多个附加资源。动态网页在设计实现时可以通过附加资源的方式向网页动态传递参数，根据不同参数值，在浏览器显示不同内容的网页。例如，通过“http://www.ccpc.edu.cn/content.jsp? newsid=9221” URL 地址可能看到一篇新闻报道，而更改 newsid=9222，看到的可能就是另外一篇内容完全不同的新闻报道。另外，通过“&”可以连接多个附加资源，如“http://www.ccpc.edu.cn/content.jsp? wbtreeid=1089&wbnewsid=9221newsid=9221”。

1.2.3　网站工作原理概述

1. 服务器端与客户端

通常来说，提供服务的一方被称为服务器端，而接受服务的一方则被称为客户端。例如，当浏览者在浏览中国刑警学院网站主页时，中国刑警学院网站主页所在的服务器就称为服务器端，而浏览者的计算机就被称为客户端。

服务器端和客户端并不是一成不变的，如果原来提供服务的服务器端用来接受其他服务器端的服务，此时该服务器将转化为客户端。如果计算机上已安装了 WWW 服务器软件，此时就可以把此计算机作为服务器，称为服务器端，浏览者可以通过网络访问到该计算机。在后面实现操作过程中，通常通过虚拟机模拟 WWW 服务器端，本机作为客户端，当然也可以把本机即当作服务器，又当作客户端。

2. 静态网站工作原理

所谓静态网站，是指在服务器端发布的是静态网页，即在网页文件里不存在程序代码，只有 HTML 标记，其文件后缀名为 htm 或 html。静态网页创建成功后，其中的内容不会再发生变化，无论何时何人访问，显示的内容都是一样的，如果要对其中的内容进行修改、添加、删除等操作，就必须到程序的源代码中进行相关操作，再重新上传到服务器上。

当用户在浏览器上输入 URL 网址，并按 Enter 键后，表明向服务器发出了一个浏览网页的请求，其实静态网站的工作原理很简单，就是一个请求和响应的模式，客户端请求

服务器端，服务器把请求的数据相应返回给客户端。当服务器收到请求后，就在其管理的主目录下查找用户所要浏览的网页，找到后将其原封不动的再发送给客户端。由客户端浏览器解释执行静态网页，为用户显示精美的网页。其原理如图 1-2 所示。

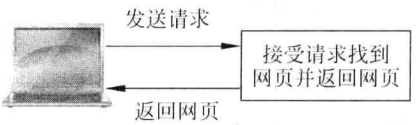

图 1-2　静态网站工作原理

3. 动态网站工作原理

动态网站是指在服务器端发布的是动态网页，在动态网页中不仅包含 HTML 标记，同时还包含实现相关功能的程序代码，该网页的后缀通常根据程序语言的不同而不同。例如，ASP 文件的后缀为 asp，而 JSP 文件的后缀则为 jsp。动态网页可以根据不同的时间、不同的浏览者而显示不同的信息。例如，常见的留言板、论坛、博客等都是应用动态网页实现的。其工作原理如图 1-3 所示。

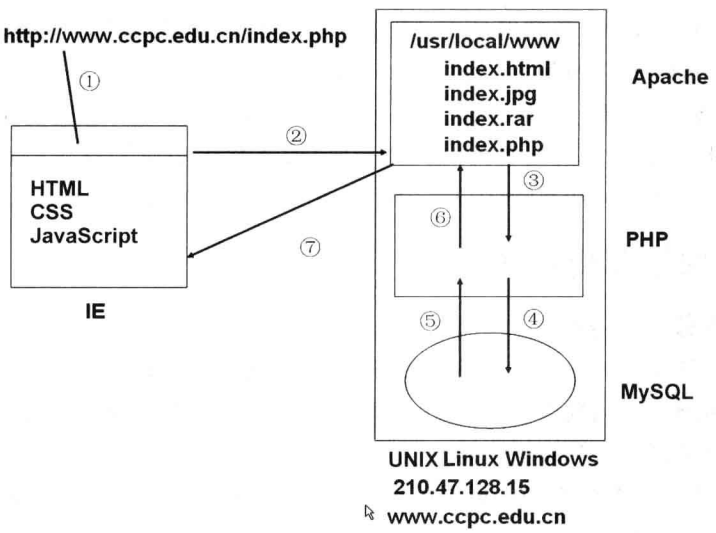

图 1-3　动态网站工作原理

例如，用户请求的是“http://www.ccpc.edu.cn/index.php”，WWW 服务器 Apache 接收到该请求后，在其所管理的主目录中查找 index.php，找到后如果仅仅将该文件原封不动地返回给客户端，这时，客户端会出现弹出下载对话框如图 1-4 所示。

因为浏览器不能解释执行 PHP 程序代码，而 Apache 本身又完成不了对 PHP 的解释工作。因此还要在 WWW 服务上还要装一个应用服务器专门解释后台程序代码的、并且在 Web 服务器里面设置针对后缀名为 PHP 的客户请求，找到后不直接发给浏览器，而是将 PHP 文件转发到 PHP 应用服务器上去解析，如果网页涉及数据存取操作，则需要在动态网页文件中嵌入 SQL(Structure Query Language)来实现，那么还必须在 WWW 服务器中安装数据库软件，或单独构建数据库服务器。以用户请求“http://www.ccpc.edu.cn/index.php”为例，在服务器端对该动态网页请求的具体处理流程为：