

计算机取证技术体系研究

戴士剑



第二届中国计算机取证技术峰会

THE 2ND CHINA COMPUTER FORENSICS CONFERENCE

2006年6月17-18日

17-18 June 2006

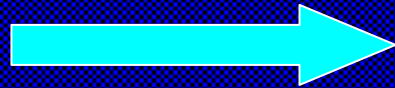
主要内容

1. **计算机取证的一般概念**
2. 计算机证据获取的一般步骤
3. 证据分析的技术体系 *
4. 展望

磁光电形式的数字信息:

传统媒介:

报纸
书刊
杂志
音像
图片
等等



易复制
易携带
易自动管理
易检索
脆弱
易泄密
易被破坏
不留痕迹
形式多样
容易嵌入、隐藏
不能直接观察.....

信息技术的高速发展，不是只改变了人类的某个行业，也不是只改变了人类的交流方式，而是完完全全的改变人们的生活。其所带来的冲击是前所未有的，涉及到人类生存的方方面面，谁也无法逃避。几乎所有的传统方式都面临着新的挑战，接受也罢，抗争也罢，最终都将融合在一起，并被毫不留情的打上信息化的烙印。

同样地，对于理性社会的社会运行体系，尤其是对法律体系的挑战，更是超出以往任何一次技术变革所带来的冲击。一切诸如国家安全问题，电力、电信基础网络安全问题，金融系统安全问题，知识产权问题，网络诈骗问题，IP电话问题、黄色泛滥问题等等，举不胜举。

要解决这些问题，既要解决法律层面的挑战，也要解决技术层面的挑战。

计算机取证技术就是打击信息技术犯罪的一项有效手段。该技术将计算机系统视为犯罪工具或现场，运用先进的技术手段，按照规程全面检查计算机系统，提取、保护并分析与计算机犯罪相关的证据，并据此提起诉讼。

确切的说, 电子证据的范畴非常广泛, 本文中讨论的计算机证据只是它的一个子集, 以下不再注明。计算机证据, 是指在计算机或计算机系统运行过程中产生的, 以其记录的内容来证明案件事实的电、磁、光等各种记录物上所存储的数据信息。

计算机证据在计算机屏幕上的表现形式具有多样性，尤其是多媒体技术的出现，更使这些证据综合了文本、图形、图像、动画、音频及视频等多种形式，这种以多媒体形式存在的计算机证据几乎涵盖了所有的传统证据类型。因此相应地，就有了电子物证、电子书证、电子视听资料、电子证人证言、电子当事人陈述、关于电子证据的鉴定结论以及电子勘验检查笔录七种类型。

以“北大、清华爆炸案”为例，案件本身并不涉及计算机系统，但却是通过计算机系统破案的。我国学者何家弘先生对电子证据做过一个断言：“就司法证明方法的历史而言，人类曾经从“神证”时代走入“人证”时代；又从“人证”时代走入“物证”时代。也许，我们即将走入另一个新的司法证明时代，即电子证据时代。”诚然，这句话并不是说，电子证据将取代一切传统证据，但信息时代，电子证据不可避免的被一些专家誉为“证据之王”。

在作为证据的共同要求方面，计算机证据与传统证据一样，必须可信、准确、完整、符合法律法规，即可为法庭所接受。但由于计算机证据所具有的特殊性，如何对其进行收集、保护、分析和展示，成了司法和计算机科学领域新的研究课题。

计算机证据与传统证据的最大区别就在于其脆弱性、易毁性、隐蔽性及非直观性。不借助一定的环境，这些证据似乎是看不见摸不着的，而传统证据如文书，可以直接观察，然后才是不容易改变，或者改变后会或多或少留下痕迹。而计算机证据，即使借助特定的环境，所看到的信息也不一定就是其本身的面目，甚至看到的根本就与其本质毫无关系，因此，计算机证据往往更难于获取，给人一种不可捉摸，尤其是无法把握，危险万分的印象。

计算机系统在相关的犯罪案例中可以扮演黑客入侵的目标、作案的工具和犯罪信息的存储器3种角色。无论作为哪种角色，计算机及其外设中都会留下大量与犯罪有关的信息，电子取证就是对计算机犯罪的证据进行获取、保存、分析和出示，其技术实质就是对计算机系统进行处理，得到相关数据，从而重建其犯罪过程。

主要内容

1. 计算机取证的一般概念
2. 计算机证据获取的一般步骤
3. 证据分析的技术体系 *
4. 展望

计算机证据的获取一般分为两大步骤，第一步是实体物理设备或软件系统的获取，即计算机系统的获取，第二步是证据分析。

物理证据获取是全部取证工作的基础，在获取物理证据时最重要的工作是保证获取的原始证据不受到任何破坏，无论在任何情况下，调查者都必须牢记以下几点：

- (1) 不改变原始记录;
- (2) 不在作为证据的计算机上执行无关的程序;
- (3) 不给犯罪者销毁证据的机会;
- (4) 详细记录所有的取证活动;
- (5) 妥善保存得到的物证。

获取物理证据后，接下来的工作就是信息发现。不同的案例对信息发现的要求是不一样的，在有些情况下，只需找到关键的文件、图片或邮件就可以了，在其他时候则可能要求重现计算机在过去工作的细节，比如入侵取证。

为了保护原始数据，除非有特殊的需要，所有的信息发现工作都是对原始证据的物理拷贝进行的，而不是直接对原始设备进行操作，并且，所有的工作都是可重复难证的。

一般情况下，取证专家还要用MD5或其他算法对原始证据上的数据做摘要，然后把原始证据和摘要信息及相关文档妥善保存。

最后，取证专家会就计算机信息发现的结果做出完整的报告，这个报告将成为打击犯罪者的依据。

由于信息系统具有复杂多变性，并且发展迅速，所以信息发现的结果在很大程度上还依赖于取证专家的经验 and 智慧。这就要求一个合格的取证人员必须对信息系统有深刻的了解，掌握计算机组成、操作系统、文件系统、分布式计算、数据库、网络体系和协议等多方面的知识。

那么，取证专家到底是如何工作的呢？除一般要求外，还需要具备什么知识？

首先，计算机取证分为两大类：一类是有准备的取证，我们这里姑且称之为有预谋的取证，这种取证的最大特点是周期较长，并具有主动性、计划性和针对性。常常是发现了蛛丝马迹后，主动采取一些技术手段，进行监视和跟踪。另一类与此相对应，完全是事后调查取证，通常具有盲目性，或只有一般线索，并不能确定获取的物理证物中一定就能找到证据。

实际的调查取证中，会涉及到各种各样的应用环境。从这个角度来讲，最主要的两种类型是网络环境和单机环境。网络环境下，信息会象空气那样弥漫到网络空间中，一方面，会使罪犯在多个地方留下犯罪痕迹，并且对其来说，更难销毁这些证据，比如服务器上的访问记录、路由器中的日志信息等，但另一方面，收集这些证据也会更困难，时效性要求更高。

除这些特征外，与单机环境相比，其另一大特点就是，网络环境下更容易采取主动取证手段，如网络监控、网络抓帧，通过服务器或代理服务器记录嫌疑人计算机所有的网络活动，或采用合法的类似于黑客的技术手段等，对重点目标进行24小时的监控等，尤其是网络抓帧技术和服务器记录，可以很容易地实现对犯罪嫌疑人网络活动的监控，而不会被发现，如图1所示。

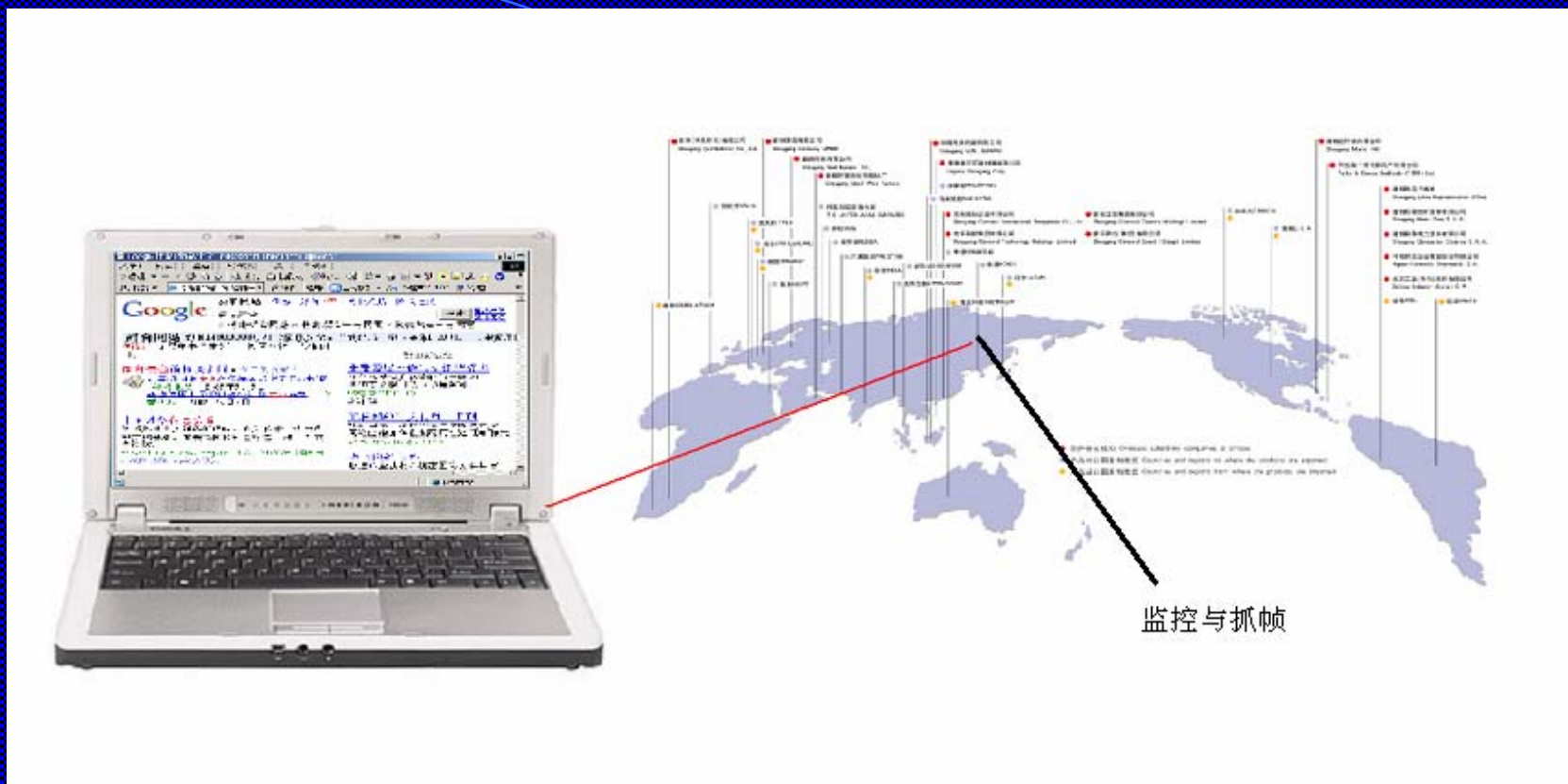


图1 主动网络取证

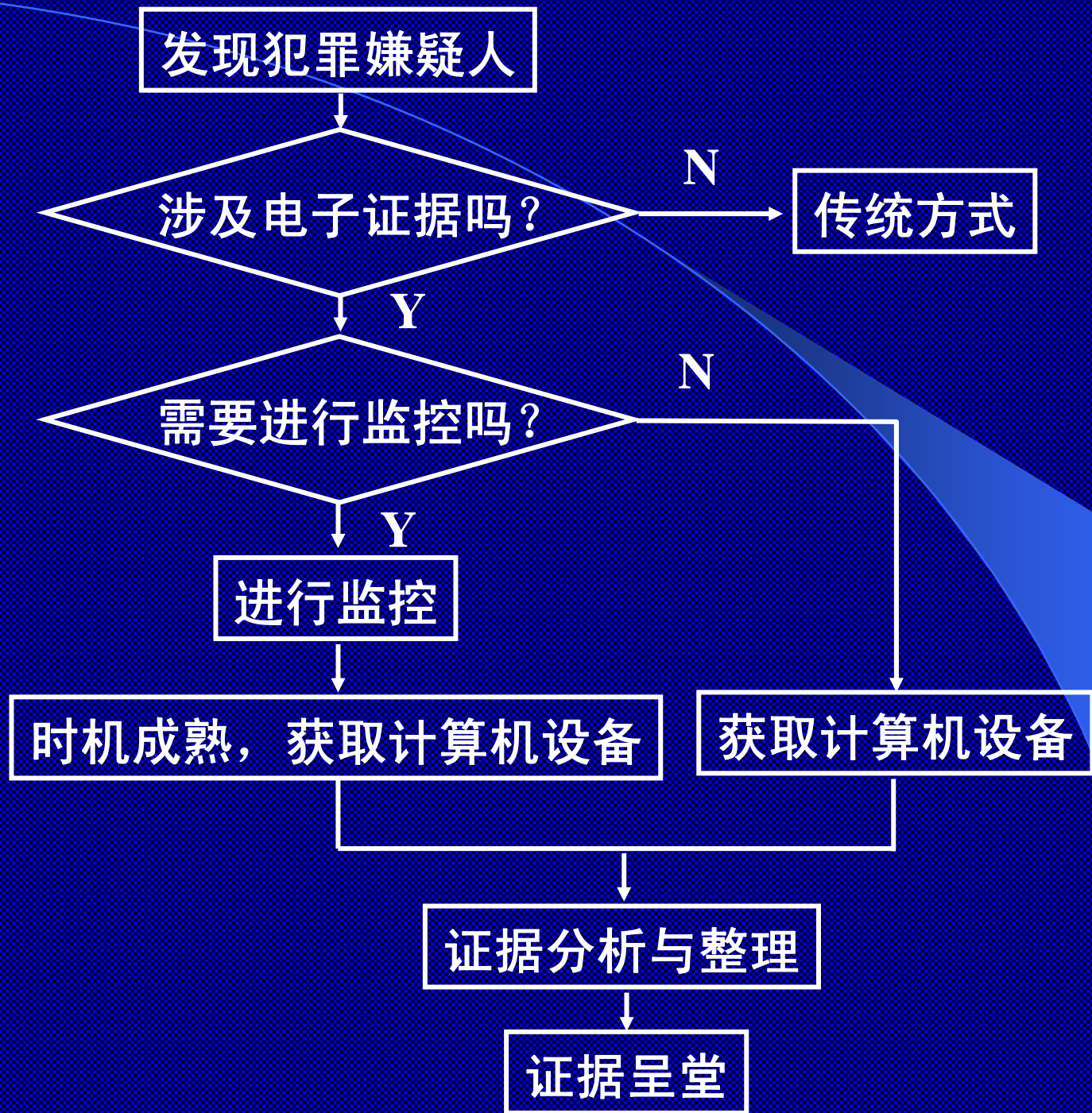
单机环境是指对不联网机器的取证调查。这类应用往往是一次性的，突击性的。比如调查时获取相应地计算机或物理存储设备，然后对其进行分析，查找相应的证据，因而更多的应用是事后取证。但也不是绝对的，也有可能在犯罪嫌疑人使用的没有联机的计算机中种下木马、间谍程序之类的东西，来收集证据，但要注意合法使用，且不要被发现。

对网络取证和单机取证做一个简单的比较，如表1所示。

表1 网络取证和单机取证比较

序号	内容	网络取证	单机取证	备注
1	信息分布	网络及本地机	本地机	
2	证据销毁	不易完全销毁	容易销毁	
3	时效性	要求较高	相对低一些	仅相对而言
4	实时监控性	可进行网络监控	一般无法进行监控	
5	信息冗余性	一般有重复记录	一般无重复记录	仅相对而言
6	是否有主动性	可主动监控	一般无法主动监控	仅相对而言

无论是网络环境下的计算机取证，还是单机环境下的计算机取证，都是一个获取证据的过程，在这个过程中，会涉及到方方面面的技术应用，一个初步的流程可以归结为图2所示。



事实上，真正的计算机取证过程是非常复杂的，需要很多的专业知识做背景，比如证据的获取，远非一般人想象的那么简单。首先，在扣押物理设备时，如果计算机处于开机状态，是立即切断电源，还是正常关机这么一个小小的细节，就非常有学问，处理不当可能就会造成证据破坏。

取证人员必须设法保存尽可能多的犯罪信息。由于犯罪的证据可能存在于系统日志、数据文件、存储器、交换区、隐藏文件、空闲的磁盘空间、打印机缓存、网络数据区和计数器、用户进程存储区、堆栈、文件缓冲区、文件系统本身等不同的位置，要收集到所有的数据是非常困难的，在关键的时候要有所取舍。

如果现场的计算机是黑客正在入侵的目标，为了防止犯罪者销毁证据文件，最佳的选择也许是马上关掉电源，实际应用中的情况会更加复杂一些。

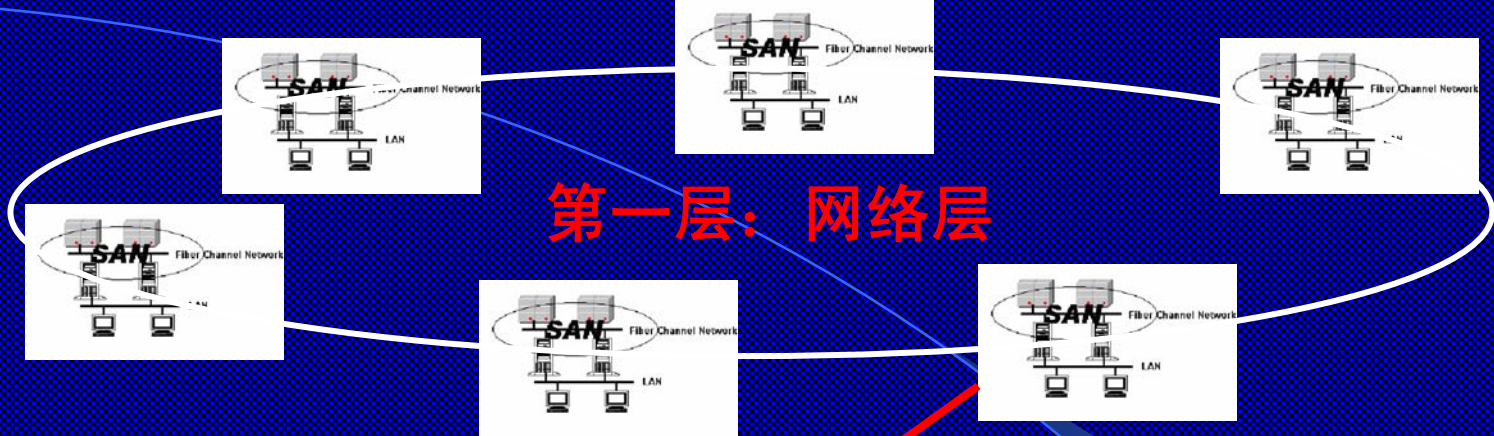
另外，存储介质和接口的多样化，也给制作原始数据的镜像工作带来了很大的挑战，再加之系统的多样性，仅仅这些一般性的操作就已经够复杂的了。

主要内容

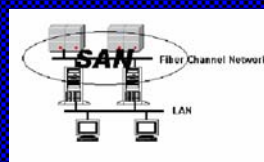
1. 计算机取证的一般概念
2. 计算机证据获取的一般步骤
3. 证据分析的技术体系 *
4. 展望

由于计算机系统的复杂性，以及数据信息的严格性，导致证据分析是一项技术性很强的复杂的工作。比如犯罪嫌疑人用于记录非法交易的WORD文档，如果被删除，在系统中就不存在了，这时，通过技术手段将该文档完整的恢复回来，并且这个过程可以无限次的重复，那么得到的信息，就是一个完整的信息，就能够得到其交易记录。

但很可能，这个文档已经部分遭到破坏，不能够完整恢复，这时就需要分析磁盘，分析该文件，尽可能多的恢复其中的内容。从这里就可以看到，取证工作的层次性很强，涉及到的技术非常多。简单看一下，除了前面提到的数据镜像的获取，还有各个层次的分析与获取，当然，这种分层并不是绝对的，并且是相关的。简要分析如下：



第一层：网络层



第二层：网络存储层



第三层：磁盘阵列层



第四层：磁盘层



第五层：文件系统层



第六层：文件层

第一层：网络层，通过网络技术获取证据。

本层次既可以应用于主动取证，也可以应用于被动取证，但典型应用为主动取证，如在犯罪嫌疑人不知觉的情况下，通过高速网络抓帧技术，得到犯罪嫌疑人所有进出网络的数据，进而重现其所有的网络活动；或者通过在路由器、服务器或代理服务器等必经地进行设置，得到其所有网络活动记录。

第二层：网络存储层，指通过网络存储层获取证据。

该层次针对存储子系统，主要包括：直接附加存储（DAS）、存储区域网络（SAN）、网络附加存储（NAS）三种形式。

DAS是Direct Attached Storage的缩写，指“直接附加存储”，将外置存储设备通过连接电缆，直接连接到一台服务器上，如图3所示：

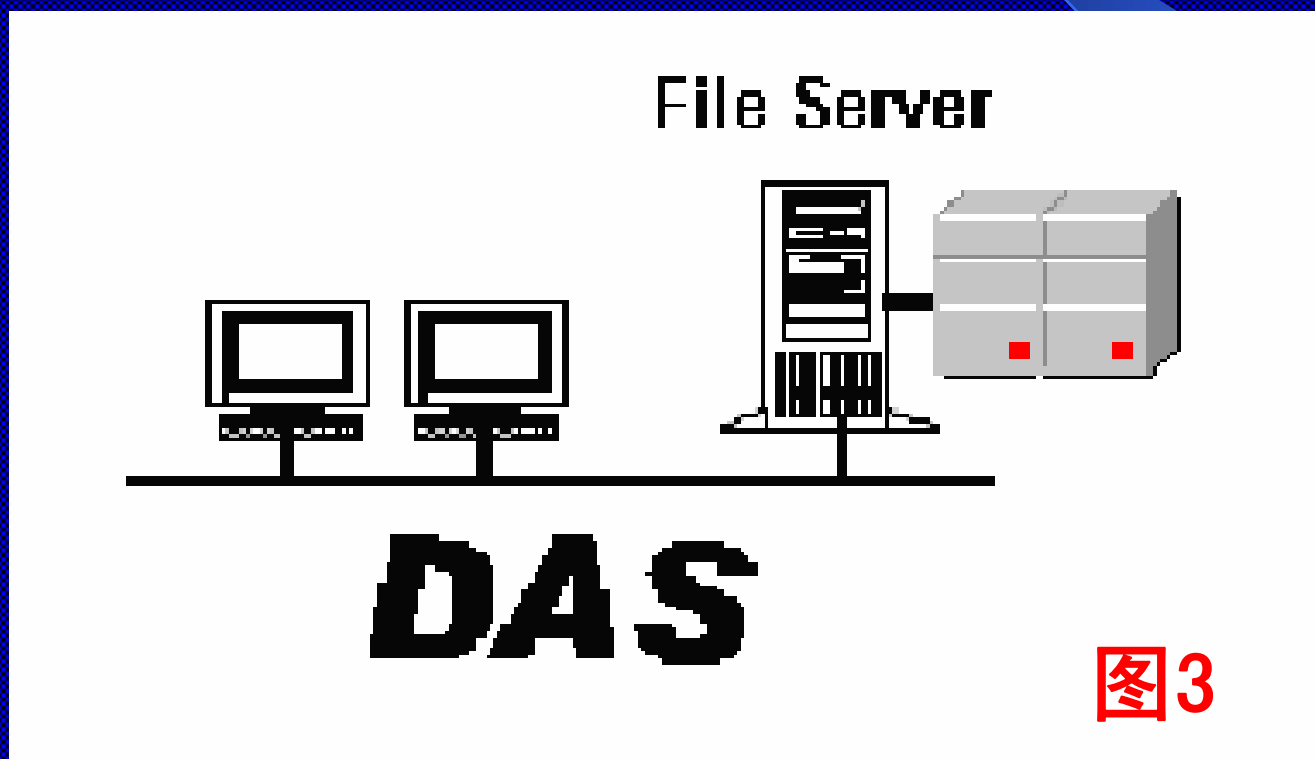


图3

采用直接附加存储方案的服务器结构如同PC机架构，外部数据存储设备采用SCSI技术或者FC（Fibre Channel）技术，直接挂载在内部总线上，数据存储是整个服务器结构的一部分，这种情况下的数据和操作系统往往是不分离的。所以，很多时候的取证工作需要将操作系统与存储系统结合起来进行分析。

NAS是英文Network Attached Storage的缩写，通常翻译为网络附加存储，其结构如图4所示：



图4

NAS作为一个网络附加存储设备，采用了信息技术中流行的嵌入式技术，使得NAS具有无人值守、高度职能、性能稳定、功能专一的特点。NAS设备内置优化的独立存储操作系统，可以实现在不同操作系统平台下的文件共享应用。

NAS设备提供RJ-45接口和单独的IP地址，可以将其直接挂接在主干网的交换机或其他局域网的Hub上，通过简单的设置（如设置机器的IP地址等）就可以在网络即插即用地使用NAS设备。

鉴于NAS的特点，其取证工作会更加复杂一些，取证工作也建立在对NAS系统访问的基础上。同时，不仅要从操作系统中提取数据，还需要从存储系统中提取数据，并且由于多机对一个NAS系统，还需要分析数据的隶属关系。

SAN是Storage Area Network的缩写，指的是“存储区域网络”，其结构如图5所示：

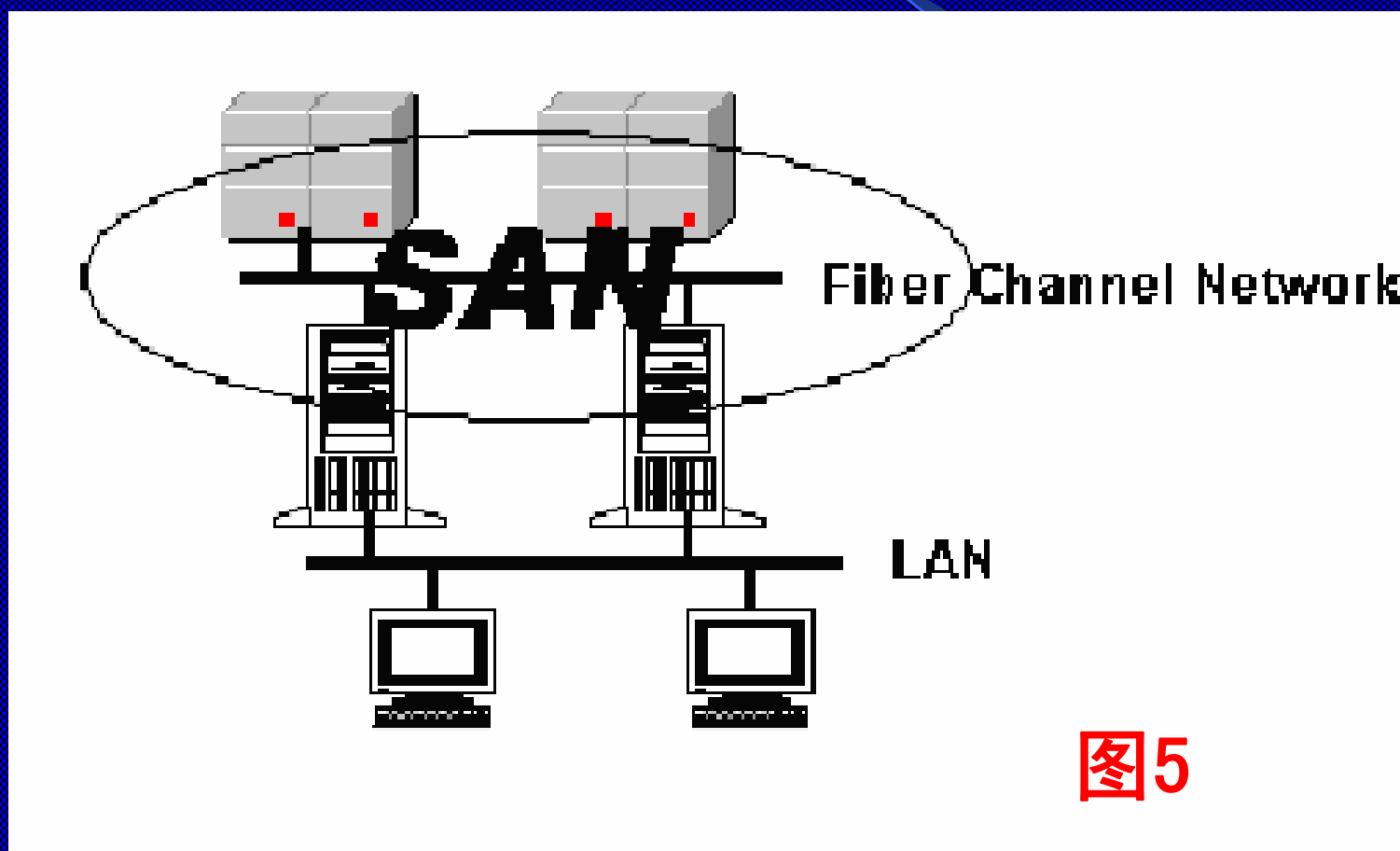


图5

SAN 采用光纤通道 Fibre Channel (FC) 技术，支持多种高级协议，其最大特性是将网络和设备的通讯协议与传输物理介质隔离开，支持多种拓扑结构，主要有：点到点 (Links)、仲裁环 (FC-AL)、交换式网络结构 (FC-XS) 等。

在SAN中，存储设备通过专用交换机连接到一群计算机上，在该网络中提供了多主机连接，允许任何服务器连接到任何存储阵列，让多主机访问存储器和主机间互相访问一样方便，这样不管数据放置在哪里，服务器都可直接存取所需的数据。

NAS和SAN最大的区别就在于NAS有文件系统和管理系统，而SAN却没有这样的系统功能，其功能仅仅停留在文件管理的下一层，即数据管理上。但NAS和SAN并不冲突，甚至常常共存于一个系统网络中。NAS通过一个公共的接口实现空间管理和资源共享，SAN仅仅为服务器存储数据提供一个专门的快速后方通道。因此，NAS的数据提取与分析更加复杂。

在本层次下搜集证据，尽量直接在本层次内解决，只有在本层次内不能解决时，才在下一层中进行解决。尤其是一些特殊的文件，更需要在这个环境下进行分析，以得到更为直接的证据。

第三层：磁盘阵列层。其实第二层的存储网络基本上都使用磁盘阵列作为基本的存储设备。在这个层次上，主要是解决阵列散架、阵列卡损坏、磁盘掉线等故障，如果要搜查的证据分布在多个磁盘上，对单个磁盘的搜索是没有任何意义的，必须重建RAID才能够得到有用的证据。

实现RAID可以采用两种方法，一种是硬RAID，用专门的控制器来完成，也就是常说的RAID卡；另一种是软RAID，用软件的方法来实现。

过去RAID一直是高端服务器才应用的设备，与高档SCSI硬盘配合使用。近来随着技术的发展和产品成本的不断下降，IDE硬盘和SATA硬盘的性能都有了很大提升，加之RAID芯片的普及，使得RAID技术也应用到了IDE硬盘和SATA硬盘上，有些主板就直接集成了RAID控制芯片，所以，取证涉及到RAID信息是必然的趋势。

RAID类型包括：RAID0、1、2、3、4、5、6、7以及一些组合方式如RAID10等，常用的RAID类型主要有RAID0、RAID1和RAID5。

RAID0是无冗余、无校验的磁盘阵列，实现RAID0至少需要两个以上硬盘，它将两个以上的硬盘合并成一块，数据按一定的规律同时分布在每块硬盘中，如图6所示。所以安全性下降，只要任何一块硬盘损坏就会丢失所有数据。

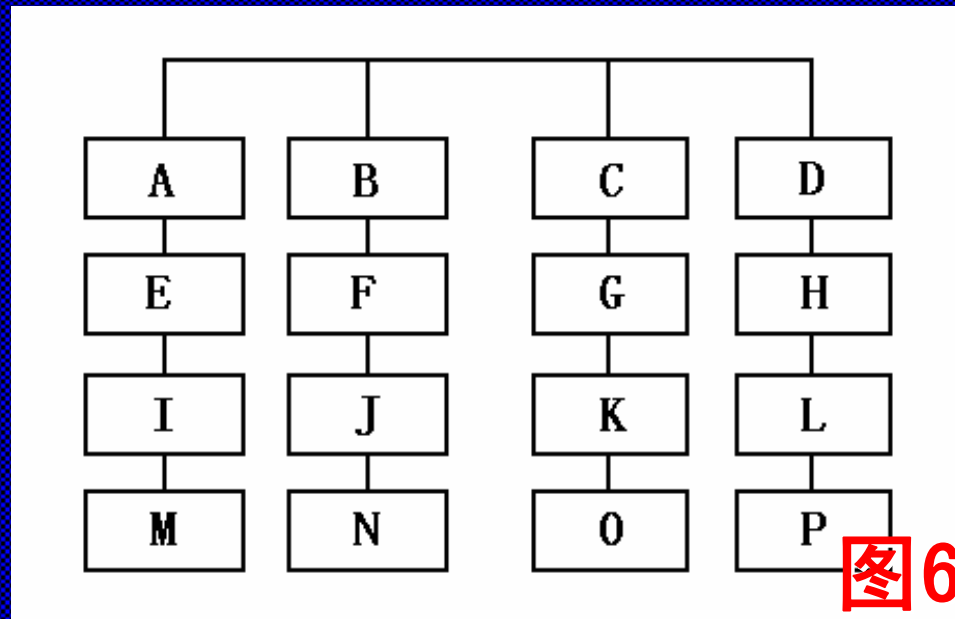


图6

RAID1又称为磁盘镜像，至少需要两个硬盘共同构建，互为镜像。如图7所示，是RAID里实现最为简单的一种方式。

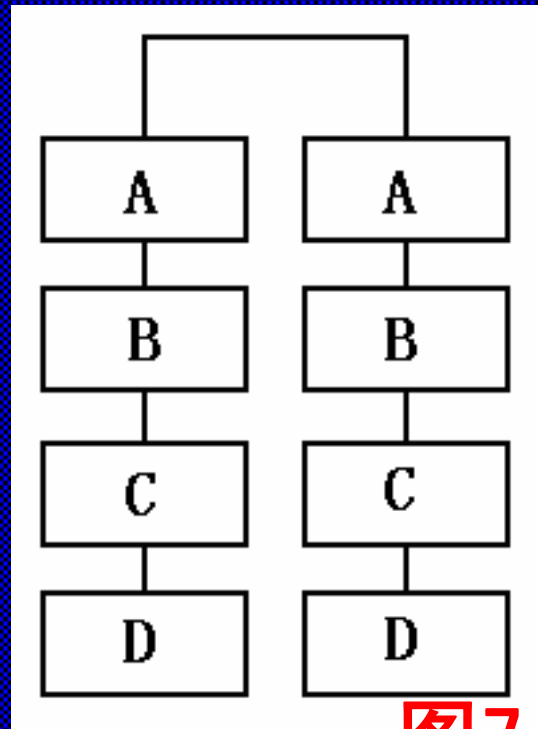
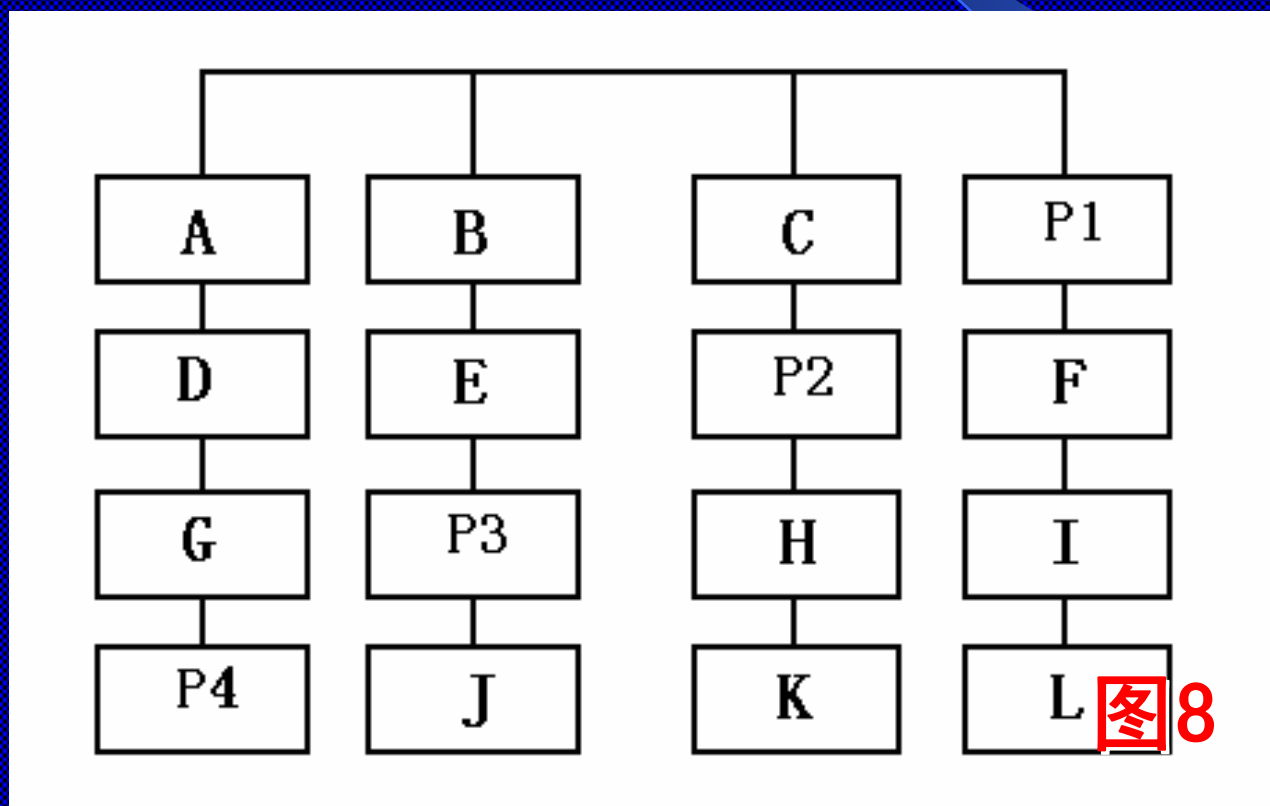


图7

RAID5的数据以块为单位分布到各个硬盘上，每一条块都有一个同样大小的校验信息，其数据分布情况如图8所示。



RAID7是优化的高速数据传送磁盘结构，这是一种新的RAID标准，其自身带有智能化实时操作系统和用于存储管理的软件工具，可完全独立于主机运行，不占用主机CPU资源，已经演变为NAS和SAN等网络存储方式。

除RAID1和极少数情况外，所有RAID中的信息，都不能孤立的进行获取，都必须建立在RAID工作正常的基础上才可以获取到需要的数据。

第四层：磁盘级。犯罪分子有时候会采取破坏设备，比如破坏计算机或硬盘的方式，以防止硬盘中的犯罪信息被提取出来，以达到销毁证据的目的。

曾经有一家企业偷税漏税很严重，当执法人员到该企业进行稽查时，该企业负责人为避免证据泄漏，将计算机的硬盘狠狠地摔在地上，导致硬盘被破坏。执法人员将摔坏的硬盘带走，在一家专业数据恢复公司进行数据恢复，数据恢复公司采用开盘更换磁头的方法成功地将该盘中的数据全部恢复出来，那家偷税漏税的企业也受到了应有的处罚。

磁盘级通常用于解决磁盘不能正常访问的处理，对上归入RAID级，对下归入文件系统级。

第五层：文件系统级。在文件系统损坏的情况下去搜索所谓的关键字显然是没有意义的，尤其是对于现在海量的大硬盘，脱离文件的关键字不仅没有意义，甚至是错误的，只有对文件系统充分了解，恢复完整的文件，才是查找证据最好的解决之道，比如，仅仅分区表损坏，将分区表重建一下，系统就完全回到正常状态，所有的文件都可以正常访问，而如果教条的用搜索关键字的方法，直接对磁盘进行搜索，显然是得不到有效数据的，效率也极其低下。

第六层：文件级。文件级包含了多种情况，如很多时候，文件系统损坏的比较严重，恢复的效果不是很理想，特别是除文本文件外，基本上各种类型的文档都有自己特定的格式，如果有损坏，就不能正常打开，很多时候犯罪分子也会有意销毁证据，

这时就需要对这些文件格式有所了解，如一个受损的WORD文档，用程序是无法正常打开显示其内容的，但可能只是文件头部分损坏，里面含有大量信息的文字并没有丢失，就可以通过技术手段，将文字信息提取出来，以获取必要的数字信息，或者一个视频资料，如果部分损坏，不能直接播放，但稍经处理，就可以播放没有损坏的部分，完全可以获取必要的视频证据。

还有就是判断文档的出处，如一张黄色图片，是某人制作的，还是传播的，等等，以及加密与解密，信息隐藏等更是取证中需要解决的问题，如将信息隐藏在图片中，不知底细的人，只能看到正常的图片，根本就无法发现和查看图片中隐藏的秘密信息。所以，文件级是差别最大、应用最复杂、需要解决问题最多的级别，包括系统本身的一些格式信息，也都可以归入到这一级。

所有这些级别基本只是针对信息流和存储系统而言的，事实上，取证工作与操作系统是密切相关，不可分割的，不仅仅是操作环境，犯罪信息本身，就有很多散布在操作系统的信息中，如注册表、虚拟内存、系统日志、缓存等处，并且还要熟悉操作系统后台会进行什么操作，以方便重建犯罪现场。

主要内容

1. 计算机取证的一般概念
2. 计算机证据获取的一般步骤
3. 证据分析的技术体系 *
4. **展望**

作为一门新兴的交叉学科，计算机取证技术还需要更多的人来投入其中进行深入研究和分析。目前的计算机取证技术尚处于发展初期，既有未解决的技术问题，更有未解决的法律问题，需要各方面的相关人员共同努力，来促进其健康、有序、有效的协调发展。技术问题主要依靠研究机构和相关系统的一线工作者共同努力来实现和解决，法律问题主要依靠法律工作者和技术研究人员、一线工作者共同探讨，并进行广泛的讨论来加以解决。

对于技术问题，随着研究的深入，可以逐步克服一些限制，如对于格式化破坏的硬盘数据，可能由于文件碎片，导致通过目前的技术手段难以得到真正有效的可用证据，但随着技术的进步，比如覆盖恢复技术的成熟，就可以完整的得到格式化以前的数据，从而解决以前的恢复和取证难题。总之，矛和盾就是不断持续发展的，随着研究的深入，一定能解决一些实际的取证问题。

对于法律问题，世界各国都有将传统法加以延伸，使之适用于电子证据领域的，有针对电子证据专门制定相应的电子法律的。我国目前正在着手制定证据法，相信随着法律体系的完善，电子证据，或者计算机证据，一定会在保护社会主义建设打击各种犯罪活动方面发挥其应有的作用。

计算机取证科学，是一门技术性非常强的边缘学科，其任何工作都必须有严格的控制程序，所得结果必须经得住考验，并可以反复验证确认，要求其操作和结果都建立在网络和文件系统的基础之上，尽量获取完整可信的信息。

个别公司出于商业利益的考虑，宣扬取证就是“拷贝+分析”的过程，这种提法本身并没有错误，但如果单纯认为分析过程就是查找关键字的过程，而没有数据恢复技术的应用，单纯靠查找关键字来确定所谓的证据，那就是绝对的不负责任的说法，就是典型的断章取义，其后果必然是制造冤假错案，是对社会主义法制的极大破坏，对此，我们坚决予以反对。

讨论与提问。

参考文献

1. 《数据恢复技术》，戴士剑，涂彦晖，张喜平，电子工业出版社，2005. 3
2. 《硬盘数据高级恢复技术》，汪中夏，刘伟，戴士剑，电子工业出版社，2006
3. 《电子证据法研究》，何家弘，刘品新，法律出版社，2002. 7
4. 《中国电子证据立法研究》，刘品新，中国人民大学出版社，2005. 5