(2019春季　课程编号：011184)

曾凡平
2019信安导论

信息安全导论

第9章 安全审计与责任认定技术

中国科学技术大学 曾凡平
billzeng@ustc.edu.cn

课程回顾：第8章 网络与系统安全防护

8.1 防火墙技术

- 防火墙的概念、特性、技术，
- 自适应代理技术，防火墙的体系结构
- 防火墙的应用与发展

8.2 入侵检测技术

- 入侵检测概述，入侵检测系统分类
- 分布式入侵检测，入侵检测技术发展趋势

8.3 “蜜罐”技术

- 概念，分类，关键机制，部署结构

8.4 应急响应技术

- 应急响应的概念，应急响应策略
- 应急事件处理流程，应急响应技术及工具

第9章 安全审计与责任认定技术

- 9.1 安全审计
  - 9.1.1 安全审计概念
  - 9.1.2 审计系统的结构
  - 9.1.3 审计的数据来源
- 9.2 数字取证
  - 9.2.1 数字取证概述
  - 9.2.2 电子证据的特点和取证基本原则
  - 9.2.3 数字取证的过程
- 9.3 数字取证关键技术和工具
  - 9.3.1 证据信息类别
  - 9.3.2 来自文件的数据
  - 9.3.3 来自操作系统的数据
  - 9.3.4 来自网络的数据
  - 9.3.5 来自应用软件的数据

9.1 安全审计

9.1.1 安全审计概念

- **所谓审计，简单地说就是记录和分析用户使用信息系统过程中的相关事件**，不仅记录谁访问了系统，而且记录系统以何种方式被使用。基于对记录的系统事件的分析，能够快速地识别问题，确定是否有攻击、攻击源自何处。因此，审计本质上是一种为事后观察、分析提供支持的机制，广泛存在于信息系统中，记录、分析、报告系统中的事件。
- **安全审计则是对系统安全的审核、稽查与计算**，即在记录一切或部分与系统安全有关活动的基础上，对其进行分析处理、评价审查，发现系统中的安全隐患，或追查出造成安全事故的原因，并做出进一步的处理。

安全审计及主要功能

- 安全审计除了能够监控来自信息系统内部和外部的用户活动，对与安全有关的活动的相关信息进行识别、记录、存储和分析，对突发事件进行报警和响应，还能通过对系统事件的记录，为事后处理提供重要依据，为网络犯罪行为及泄密行为提供取证基础。同时，通过对安全事件的不断积累并且加以分析，能有选择性和针对性地对其中的对象进行审计跟踪，即事后分析及追查取证，以保证系统的安全。
- 安全审计的主要功能包括：安全审计自动响应、安全审计数据生成、安全审计分析、安全审计浏览、安全审计事件存储、安全审计事件选择等。

1)安全审计自动响应

- 安全审计自动响应是指当审计系统检测出一个安全违规事件（或者是潜在的安全攻击）时做出的响应。它是管理审计事件的需要，这些需要包括报警甚至阻断。根据审计事件的不同系统将做出不同的响应。

2)安全审计数据生成

- 该功能规定了对与安全相关的事件进行记录，包括鉴别审计层次、列举可被审计的事件类型，以及鉴别由各种审计记录类型提供的相关审计信息的最小集合。系统可定义可审计事件清单，每个可审计事件对应于某个事件级别。
- 每条审计记录至少应包含以下信息：事件发生的时间、事件类型、主题标识、执行结果、引起事件的用户标识以及对每一个审计事件与该事件有关的审计信息。

3)安全审计分析

- 该功能定义了分析系统活动和审计数据，来寻找可能的或真正的安全违规操作。它可以用于入侵检测或对安全违规的自动响应。当一个审计事件集出现或累计出现一定次数时，可以确定一个违规的发生，并执行审计分析。事件的集合能够由经授权的用户进行增加、修改或删除操作。

4)安全审计浏览

- 该功能主要是指经过授权的管理人员对于审计记录的访问和浏览。审计系统需要提供审计浏览的工具。通常，审计系统对审计数据的浏览有授权控制，审计记录只能被授权的用户浏览，并且对于审计数据也是有选择地浏览。有些审计系统提供数据解释和条件搜寻等功能，帮助管理员方便地浏览审计记录。

5)安全审计事件存储

- 该功能主要是指对审计记录的维护，如何保护审计、如何保证审计记录的有效性，以及如何防止审计数据的丢失。审计系统需要对审计记录、审计数据进行严密保护，防止未授权的修改，还需要考虑在极端情况下保证审计数据的有效性，如：存储介质失效、审计系统受到攻击等。审计系统在审计事件存储方面遇到的问题通常是磁盘空间用尽。单纯采用覆盖旧记录的方法是不够的，审计系统应当能够在审计存储发生故障或者在审计存储即将用尽时采取相应的动作。

6)安全审计事件选择

- 该功能是指管理员可以选择接受审计的事件。一个系统通常不可能记录和分析所有的事件，因为选择过多的事件将无法实时处理和存储，所以安全审计事件选择的功能可以减少系统开销，提高审计的效率。此外，因为不同场合的需求不同，所以需要为特定场合配置特定的审计事件选择。审计系统应能够维护、检查或修改审计事件的集合，能够选择对哪些安全属性进行审计，例如，与目标标识、用户标识、主体标识或事件类型有关的属性。

9.1.2 审计系统的结构

1. 集中式结构

- 集中式的审计系统集中地收集和分析来自于多台主机的源审计记录，所有事件信息均要传送到中央处理机上，分析并做出响应。

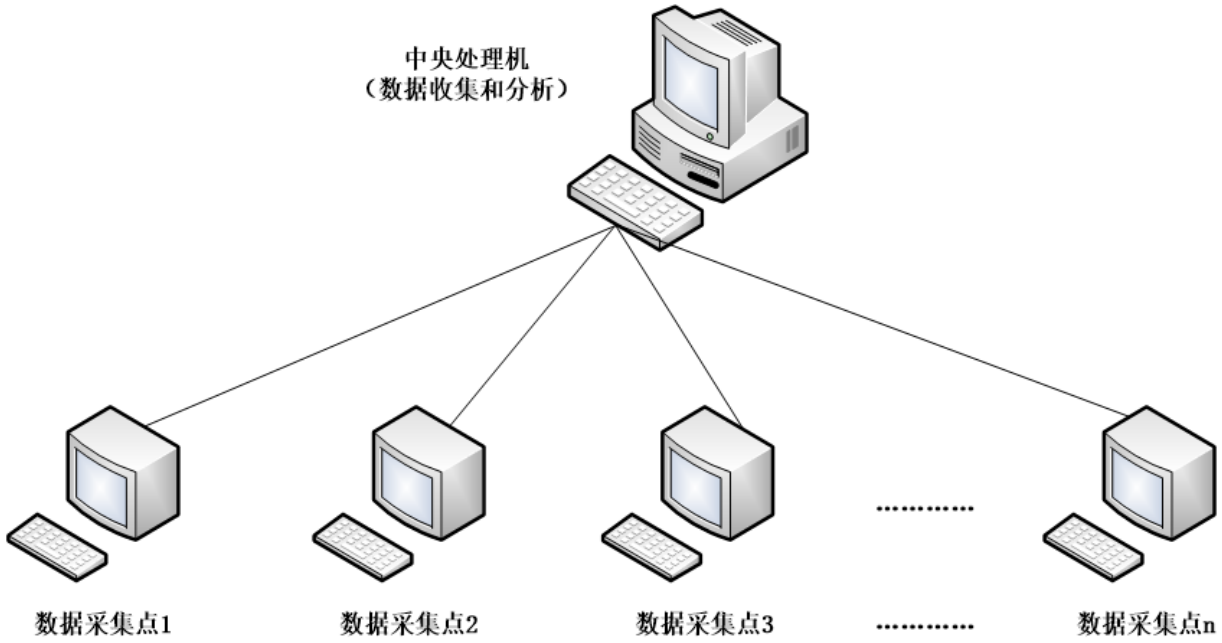


图9-1 集中式检测结构

集中式安全审计结构面临诸多挑战

- (1)集中式的审计机制中，所有的事件信息分析工作都由中央处理机完成，其CPU、I/O和网络通信的负担非常重，并且不能很好地适应大量用户的增容。
- (2)集中式的审计机制不能容括各种空间上分布的组件（如路由器、过滤器、DNS、防火墙等）及公共服务。
- (3)集中式结构存在单点失效、可扩展性有限、难以重新配置、增加功能困难等问题。
- (4)系统自适应能力差，不能根据环境变化而进行自动配置。通常配置的改变和增加是通过编辑配置文件来实现的，往往需要重新启动系统以使配置生效。

2. 分布式结构

- 分布式系统结构的安全审计任务由分布于网络各处的审计单元协作完成，这些单元还能在更高层次结构上进一步扩展，从而能够适应网络规模的扩大。

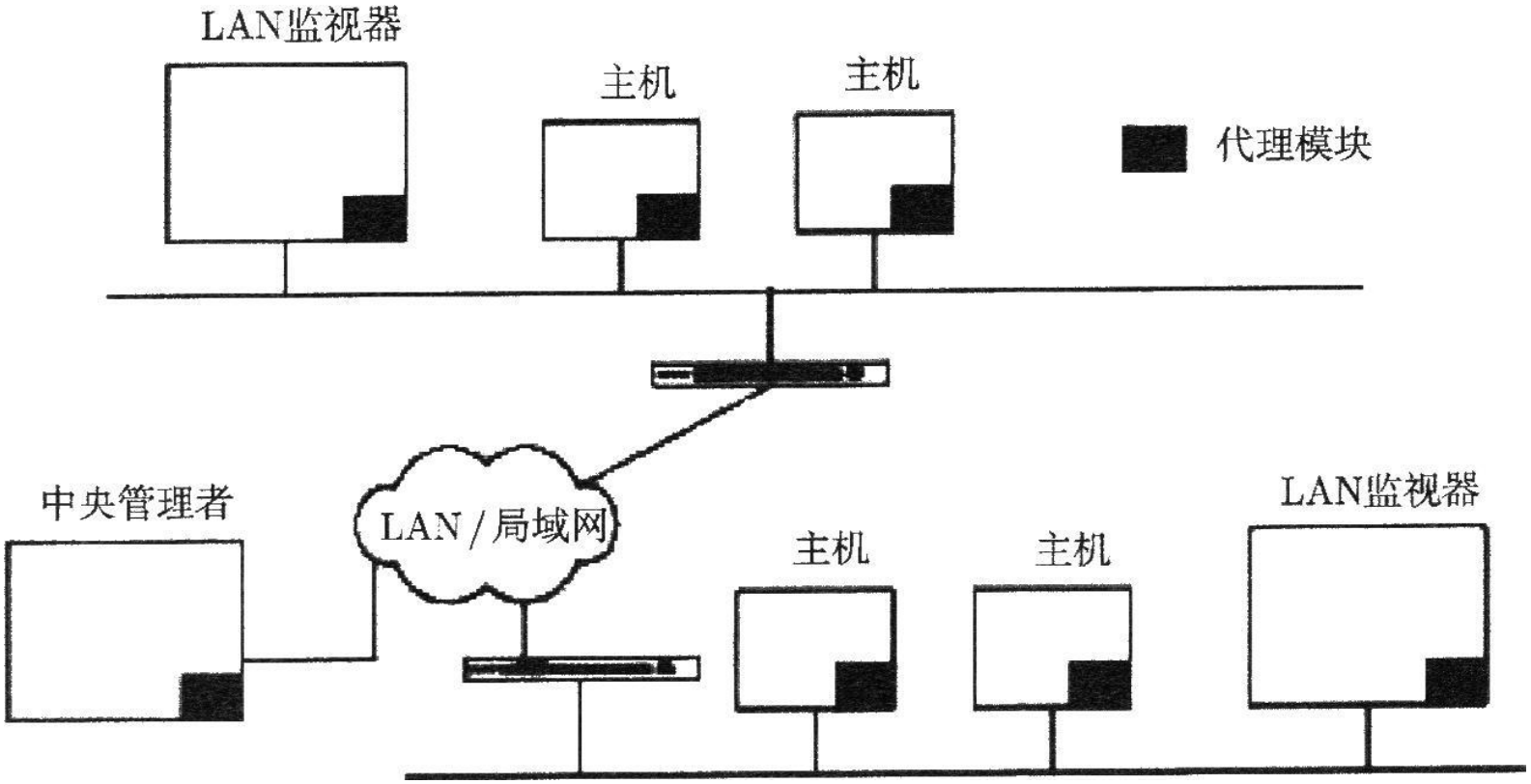
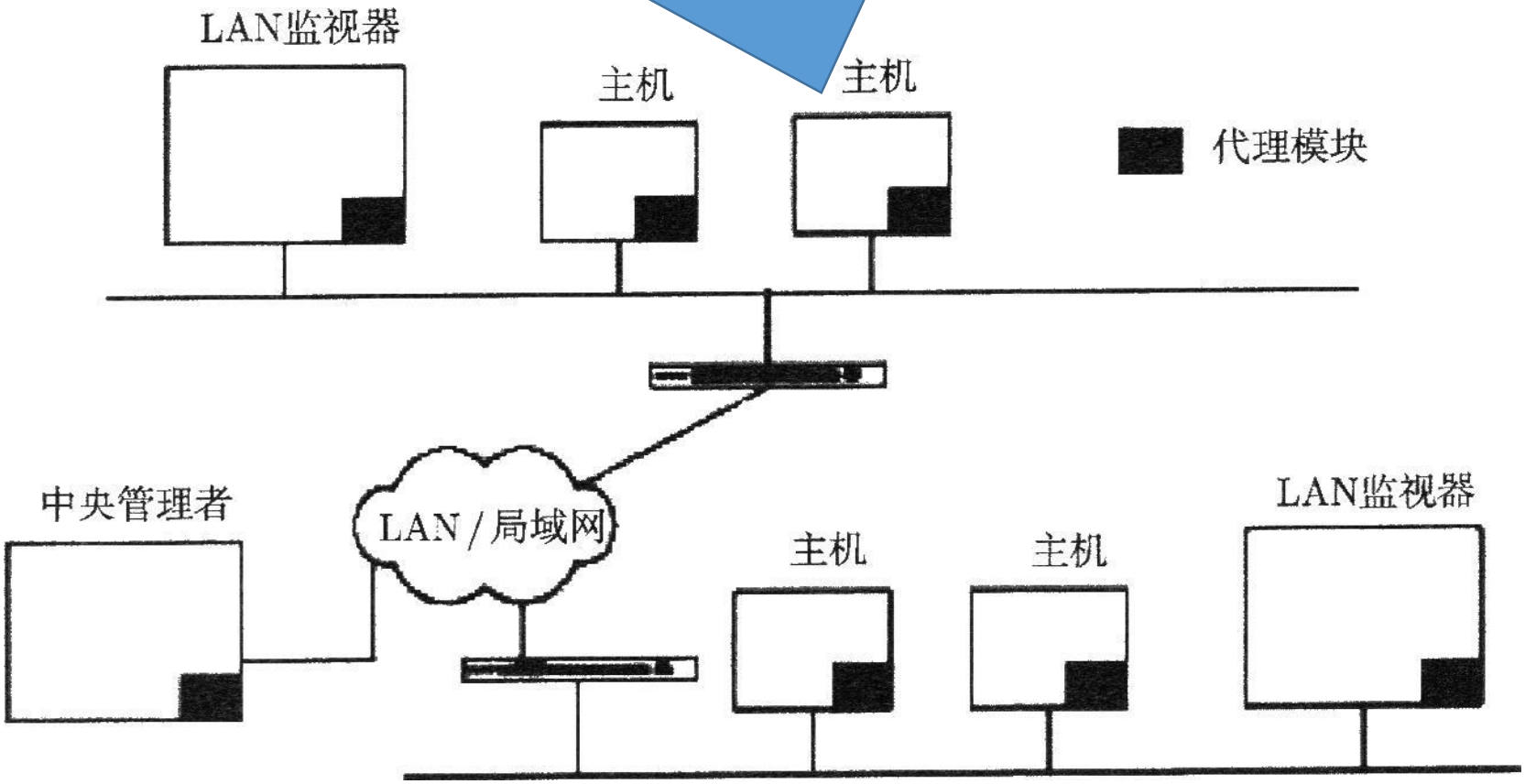
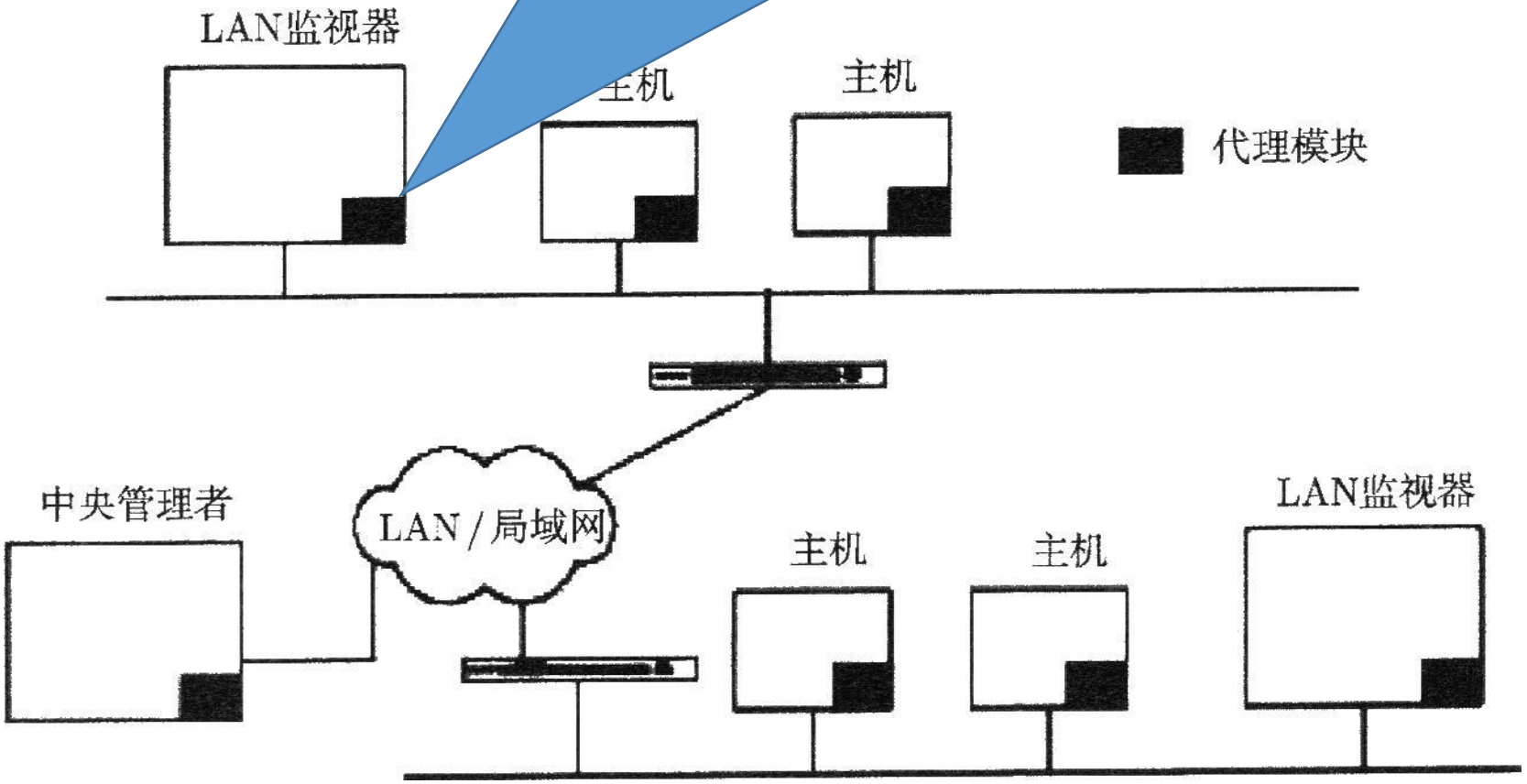


图9-2 典型分布式安全审计系统结构

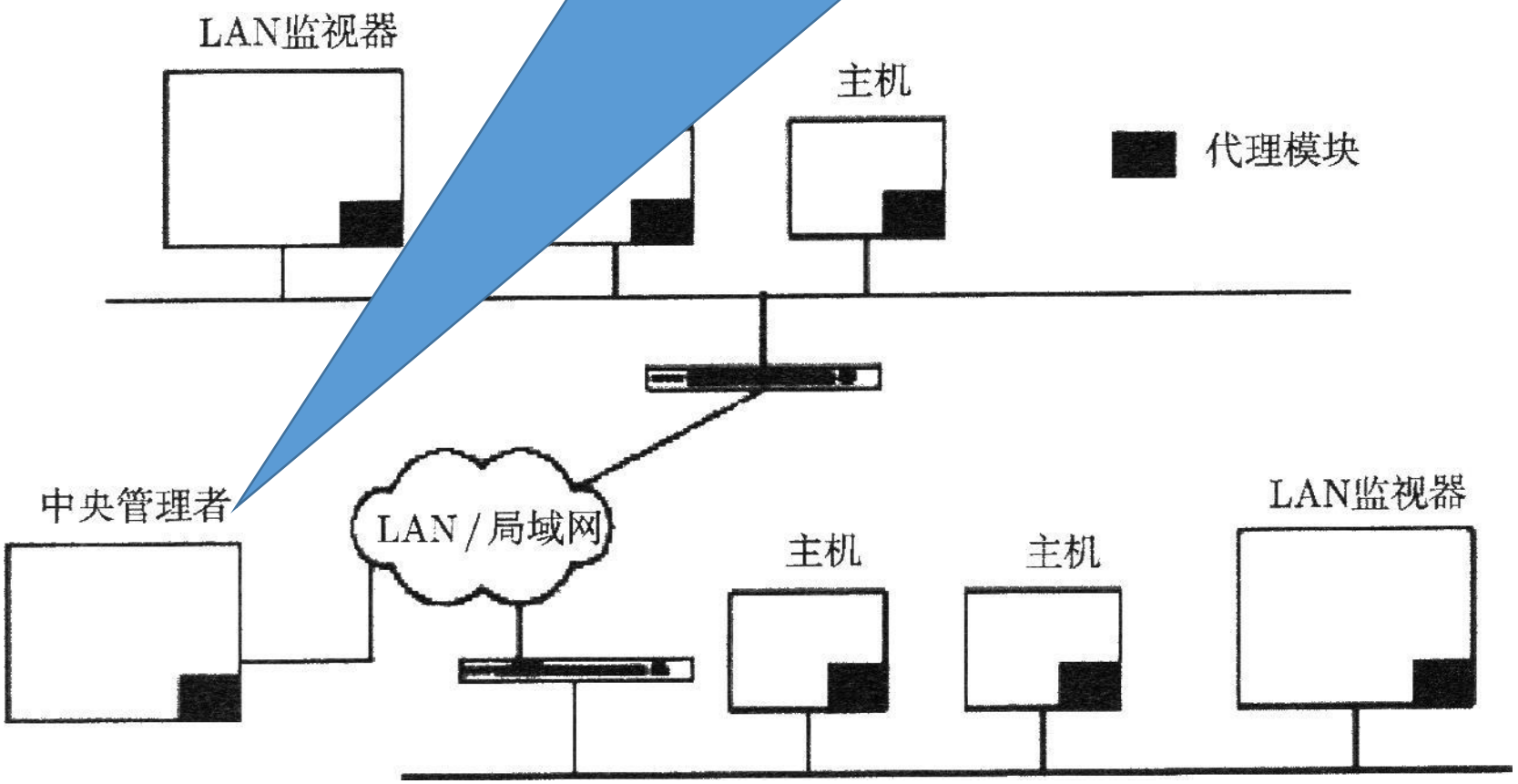
(1)主机代理模块。在受监视系统中，作为后台进程运行的审计信息收集模块。主要目的是收集主机上与安全相关的事件信息，并将数据传送给中央管理者。



(2)局域网监视器代理模块。主要分析局域网上的通信信息，并根据需要将结果报告给中央管理者。



(3)中央管理者模块。接收包括来自局域网监视器和主机代理的数据和报告，控制整个系统的通信信息，对接收到的数据进行分析。



分布式系统结构的工作过程

- 在分布式系统结构中，主机代理模块截获审计收集系统生成的审计记录，应用过滤器去掉与安全无关的记录，然后将这些记录转化成一种标准格式以实现互操作。然后，代理中的分析模块分析记录，并与该用户的历史映像相比较，当检测出异常时，向中央管理者报警。局域网监视器代理审计主机与主机之间的连接以及使用的服务和通信量的大小，以查出显著的事件，如网络负载的突然改变、安全相关服务的使用等。

- 分布式系统结构可以从单独的安全审计系统扩展成能够关联许多站点和网络行为的审计系统。

分布式系统结构的优点

- (1)扩展能力强。通过扩展审计单元来实现网络安全范围的扩张。
- (2)容错能力强。分布式的独立结构解决了单点失效问题：
- (3)兼容性强。既可包含基于主机的审计，又可含有基于网络的审计，超越了传统审计模型的界限。
- (4)适应性强。当网络和主机状态改变时，如升级或重构，分布式审计系统可以容易地作相应修改。

在设计分布式系统时要考虑以下几个主要问题

- (1)分布式检测系统可能需要采用不同的审计收集系统，而不同的审计收集系统采用不同的记录格式。因而需要处理不同的格式以实现不同节点间互操作。
- (2)网络上的一个或多个节点将作为对来自网络中各系统数据的收集和分析点。因此，原始的审计数据或者总结数据必须通过网络进行传输，因而存在保证这些数据完整性和机密性的需求。保证完整性是为了防止入侵者通过修改传输的审计信息来掩盖其行为，保证机密性是因为传输的审计信息可能是有价值的。
- (3)分析中心设置问题：可以采用集中或分散的结构。对于集中式的结构，存在对所有审计数据进行收集和分析的中央点，这方便了关联输入报告的工作，但造成了潜在的瓶颈和单点故障。对于分散式的结构，存在多个分析中心，但是它们必须能够协调各自的行为并交换信息。

9.1.3 审计的数据来源

- 大体上，安全审计系统的数据源可以分为三类：基于主机、基于网络和其他途径。

1. 基于主机的数据源

- 20世纪80年代之前，网络没有普及，审计主要针对单机系统，环境相对简单。安全审计的目的主要是检查对系统的可疑操作，防止非法入侵，并分析操作系统的系统记录和安全记录，发现可疑操作或者非法操作，修补系统的漏洞。
- 基于主机的数据源有以下四类：操作系统日志、系统日志、应用日志和基于目标的信息。

1)操作系统日志

- **操作系统日志主要由三个元素来描述，主体(subject)、客体(object)和行为(action)。**在操作系统日志中，任何一种事件都可以表示为主体对客体进行的操作。行为主要是系统服务和上层的程序应用行为，主体主要是指用户或者代表用户的操作行为，客体主要是受保护的系统资源。
- 操作系统日志的典型例子如Windows NT的事件日志(event log)，它从三个方面收集系统事件：操作系统事件、安全事件和应用程序事件。每种类型的事件使用特殊的格式记录在单独的日志文件中，通常只可以通过操作系统提供的事件查看器(event viewer)读取。

操作系统日志是首选数据源

- 操作系统日志是基于主机的审计系统或入侵检测系统的首选数据源，原因在于：
- (1)操作系统本身为审计系统或入侵检测系统及其产生的审计记录提供了实质性的保护，提高了数据源的可信度。
- (2)审计系统或入侵检测系统从操作系统的层次上获取系统信息，没有经过高层的抽象，因此，可以得到系统事件的细节，易于实现精确的模式匹配。
- (3)如果入侵者试图通过插入伪造的审计记录来达到破坏审计信息的目的，这种低层次的系统审计数据也使得入侵者的行动变得更为困难。

2)系统日志

- 系统日志是反映各种系统事件和设置的文件。UNIX系统为系统应用提供了通用的审计服务syslog，用于产生和更新事件日志。syslog在系统应用提供的文本串形式的信息前面添加应用运行时的系统名和时间戳信息，然后进行本地或远程归档。
- 系统日志的安全性较差，原因有两点：一是，产生系统日志的软件通常作为应用程序运行，容易遭到破坏；二是，系统日志通常以文本方式保存，且保存的目录一般也未受到保护。但是，由于syslog使用简单，许多系统应用和网络服务，如login、sendmail、NFS等，都使用它作为日志数据。

3)应用日志

- 目前的系统越来越趋向于面向对象和分布式结构，要想在单一操作系统层次上获取整个系统的完整信息，已经不太可能了。而应用日志通常代表了系统活动的用户级抽象信息，相对于系统的安全数据来说，去除了大量的冗余信息，更易于管理员浏览和理解。实际上，在微软的Windows NT系统中，许多原来由操作系统进行记录的事件日志已经移植到了应用级层次上，WWW服务器的日志信息是最为常见的应用级数据源。当前，主流的WWW服务器都提供访问日志机制。
- 应用日志通常具有某种特定的格式。如Apache日志支持普通记录格式(common log format)和组合记录格式(combined log format)，微软公司的IIS允许管理员选择将日志信息记录到文本文件或是ODBC(open database connectivity)数据库中，此外，IIS还支持W3C扩充日志格式和NCSA通用日志格式等。

4)基于目标的信息

- 有时，人们希望只对系统的部分活动或部分资源进行审计，由此产生了面向目标的安全审计。其方法为：评估出系统中关键的或是有特殊价值的对象，针对每一个对象制定信息收集和监视机制，该对象即为审计的目标。对于审计目标的每一次状态转变，与系统的安全策略进行比较，所出现的任何异常都进行记录或响应。
- 最常见的基于目标的审计技术是完整性校验。其审计对象多为文件，采用消息摘要算法，计算需要保护的系统对象（如关键文件）的校验值，并存储在安全区域，周期性地对目标进行检查，可以发现目标是否被改变，从而提供一定级别的保护。最具代表性的例子是Tripware。

2．基于网络的数据源

- 网络数据是目前的网络安全审计系统和商业入侵检测系统最为通用的信息来源。其基本原理是，当网络数据流在网段中传播时，采用特殊的数据采集技术，收取网络中传输的数据，作为安全审计系统的数据源。

1)网络数据的获取

- 以太网网络适配器（网卡）通常有正常模式(normal mode)和混杂模式(promiscuous mode) 。正常模式下，主机仅处理以本机为目标的数据帧；混杂模式下，网络适配器可以接收本网段内传输的所有数据帧，无论这些数据帧的目的地址是否为本机。
- 基于网络的安全审计系统必须利用网络适配器的混杂模式，获得经过本网段的所有数据信息，从而实现获取网络数据的功能。

2)采用网络数据源的优势

- (1)由于网络嗅探器(sniffer)所做的工作仅仅是从网络中读取传输的数据帧，因此，对受保护系统的性能影响很小或几乎没有，并且无须改动原先的系统和网络结构。
- (2)网络嗅探器对网络中的用户是透明的，降低了嗅探器本身遭受入侵者攻击的可能性。
- (3)网络监听相对于基于主机的安全审计，更容易检测到某些基于网络协议的攻击方法。典型的是通过向目标主机发送畸形的或大量的数据帧从而造成拒绝服务攻击(denial of service，DoS)的方法。
- (4)网络监听可针对一个网络的数据进行分析，与受保护主机的操作系统无关。与之相比，基于主机的IDS必须首先保证操作系统的正常工作，并且需要针对不同的操作系统开发不同的版本。

3．其他途径的数据源

- 除了以上介绍的几种数据源外，其他途径的数据源还有很多。比如网络系统中的设备活动日志、带外数据源等。
- **路由器、交换机等网络设备产生的活动日志，通常遵循RFC3164(the BSD syslog pro tocol)规定的日志格式，可以通过syslog实现方便的转发和处理**。一个典型的syslog记录包括生成该记录的进程名字、文本信息、设备和优先级范围等。对防火墙、IDS等安全设备所产生的活动日志，其格式则没有统一标准，目前国内多数防火墙支持WTELF(web trends enhanced log format)日志格式，而多数IDS的日志兼容Snort产生的日志格式。
- **带外数据源则是指通过人为的、非系统的方式获得的信息**。例如，手工记录系统环境中发生的事件，包括硬件错误、系统掉电、系统崩溃、自然灾难等。这些信息对于事后的分析可能起到重要作用。

9.2.1 数字取证概述

- 数字取证(digital forensics，以下简称取证)，也被称为计算机及网络取证、计算机取证(computer forensics)，就是应用计算机、通信等相关技术，发现、收集、检查、分析数据，同时保护信息的完整性，并维持严格的数据保管链。
- 数字取证的作用，就是通过调查可疑的计算机和网络系统，收集和保存证据，重建事件，评估事件的状态，获得证据，从而进行犯罪调查或者响应一个计算机安全紧急事件。
- 具体地，数字取证的作用包括以下几点：

数字取证的作用:打击违法犯罪的作用

- (1)发现和归档证据和线索。
- (2)固定其他途径发现的证据。
- (3)帮助揭示事件模型。
- (4)关联攻击和受害的计算机。
- (5)展现端到端侵害事件的路径，不管侵害是否已遂。
- (6)提取隐藏、删除或者其他不能直接得到的数据。
- 以上主要是针对打击违法犯罪的作用。除此以外，数字取证的作用还包括以下方面：

- (1)排除故障：取证工具和技术可以用来排除计算机和网络中的故障。如：发现和定位不正确的网络配置、应用程序的功能性问题，记录和审视主机的设置和运行等。
- (2)日志监控：实时取证过程中，需要监视各种日志文件，分析和关联各种系统的日志记录，从而帮助在应急响应中的事件处理、识别违反安全策略的事件、实施审计等。
- (3)数据恢复：从系统中恢复丢失的数据，包括偶然和故意删除的数据，以及由于其他原因被修改的数据。能够恢复多少数据，取决于具体的情况。
- (4)数据提取：帮助从加密的、隐含的、分散的文件及系统区域提取和还原数据。例如，解密文件和系统，提取非正常读、写的数据，提取电子邮件、聊天记录、Internet浏览记录等。
- (5)完善策略：调查用户违反安全策略的行为。例如，暴露敏感信息、关闭某些用于审计的进程等。同时，针对取证的结果，提出完善安全策略的方法，提高网络的安全性能。

1．数字取证的分类

- 按照不同的分类方法，可以将数字取证分成不同类型：
- **(1)主机取证和网络取证**：按照是否调查和涉及网络数据流分类，前者主要针对主机及其外部设备，后者针对网络数据和周界网络。
- **(2)事后取证和实时取证**：指取证调查的时间是在事件发生后还是在事件发生中。
- **(3)司法取证和非司法取证**：指取证调查得到的证据是否进入司法程序。前者的取证在步骤和证据方面有特殊和严格的要求，后者的取证结果只在企业或者组织内部使用。

2．数字取证的数据媒介

常见的数据媒介有以下方面：

- ►标准的计算机系统：桌面机、便携机、服务器。
- ►网络设备：防火墙、路由器。
- ►外部设备：打印机、扫描仪。
- ►存储设备：移动硬盘、软盘、光盘、U盘、备份磁带、ZIP盘、各种存储卡，如SD卡、CF卡、记忆棒等。
- ►消费电子产品：手机、PDA、数码相机、数码摄像机、MP3、MP5等。

9.2.2 电子证据的特点和取证基本原则

- 取证的直接目的，就是要得到说明或验证某个事件的证据。计算机取证得到的证据称为电子证据(electric evidence)或数字证据(digital evidence)。不是所有的调查数据都可以作为证据，证据必须满足**两个根本属性**。

① **可接受性(admissibility)**：在技术上或者法律上可以接受。

② **完整性(integrity)**：数据是真实、可靠的。

- **真实性(authenticity)**是指数据来源于它本来的地方并且没有被污染；**可靠性(reliability)**指的是数据所说明的事实是可信的和一致的，需要将数据与其结果关联。

电子证据的特点

- **(1)数字性。**计算机证据的物质载体是电子元器件和磁性材料等。从物理表示上，数据的变化只是集成电路的电子矩阵正负电平或磁性材料磁体发生了变化。获取这些行为的证据需要特殊的手段，电子与其他证据的获取是完全不同的。
- **(2)技术性。**计算机证据的产生、储存和传输及其采集、分析和判断都必须借助于计算机科学中的计算技术、存储技术、网络通信技术等。
- **(3)脆弱性。**计算机系统对数据的处理环节多、使用的技术和设备复杂、处理速度越来越快，数据的修改可以在瞬间完成，并且可以不留痕迹，从而使得电子证据具有脆弱和不可靠的特性。

电子证据的特点

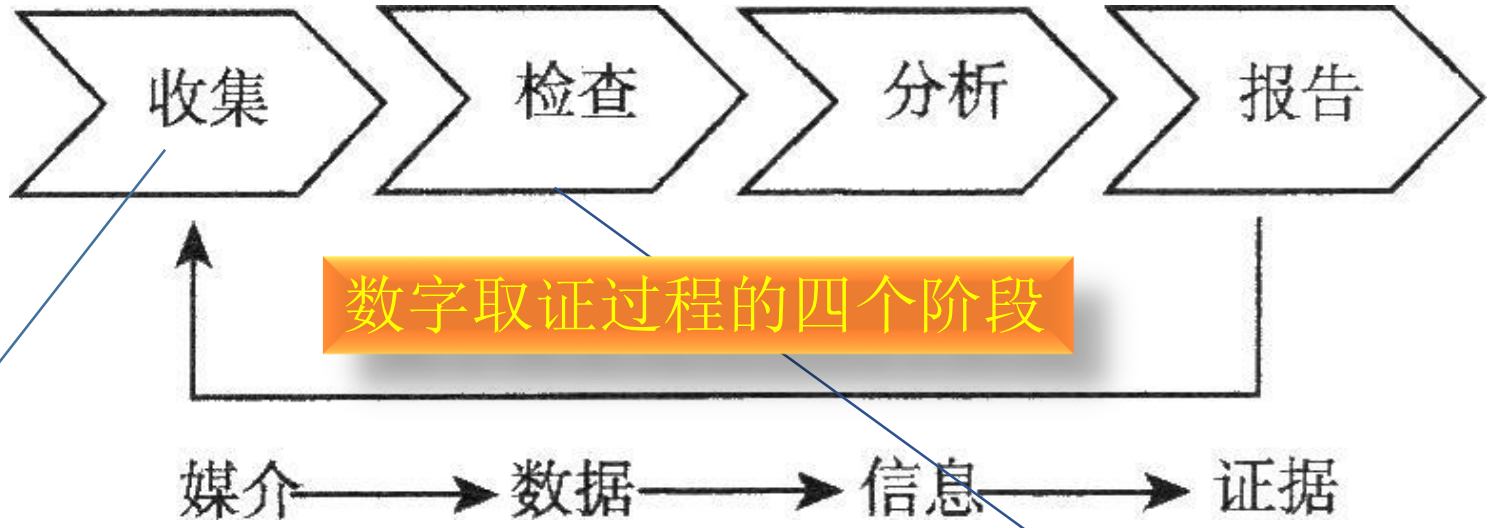
- **(4)多态性**。是指电子证据的表现形式是多种多样的，即不同形态输出材料的证明力都来源于同一计算机存储的信息本身。虽然不同的证据表现形式并不能说明其在审查判断上有根本的区别，但是不同形态证据材料的审查规则是不同的。
- **(5)人机交互性**。是指电子证据的形成，在不同的环节上有不同计算机操作人员的参与，并且会对电子证据施加不同的影响，这种影响的层次和程度与操作人员的工作性质有关。所以，可能出现的问题也就存在于人、机两个方面。为了保证证据的可靠性和真实性，应该从技术和管理上严格控制人机系统。
- **(6)复合性**。是指当证据以某种形式表现时，往往就具有了这种表现形式的证据特征。计算机证据的表现形式是多种多样的，可以是打印在纸上的文字、在显示器上输出的视频、图像、文字，也可以是声音设备输出的声音。

计算机取证应该遵循的原则

- **(1)及时性原则**：要求计算机证据的获取有一定的时效性。
- **(2)取证过程合法性原则**：要求计算机取证过程必须按照法律的规定公开进行，得到第一手具有证明力的公正的证据信息。
- **(3)多备份原则**：对于含有电子证据的媒介至少应制作两个副本，原始媒介应妥善保管，副本用于取证人员进行证据的提取和分析。
- **(4)环境安全原则**：取证过程应该在安全环境中进行，以各随时重组、试验或展示。存储证据的媒介应远离高磁场、高温、灰尘、积压、潮湿、腐蚀性化学试剂等。在包装计算机设备和元器件时，尽量使用纸袋等不易产生静电的材料，以防止静电消磁。
- **(5)严格管理过程的原则**：含有计算机证据媒介的移交、保管、开封、拆卸的过程必须由应急响应取证人员和保管人员共同完成，每一个环节都必须检查其真实性和完整性，并拍照和制作详细的笔录，由行为人共同签名，从而形成完整的证据链。

9.2.3 数字取证的过程

图9-3

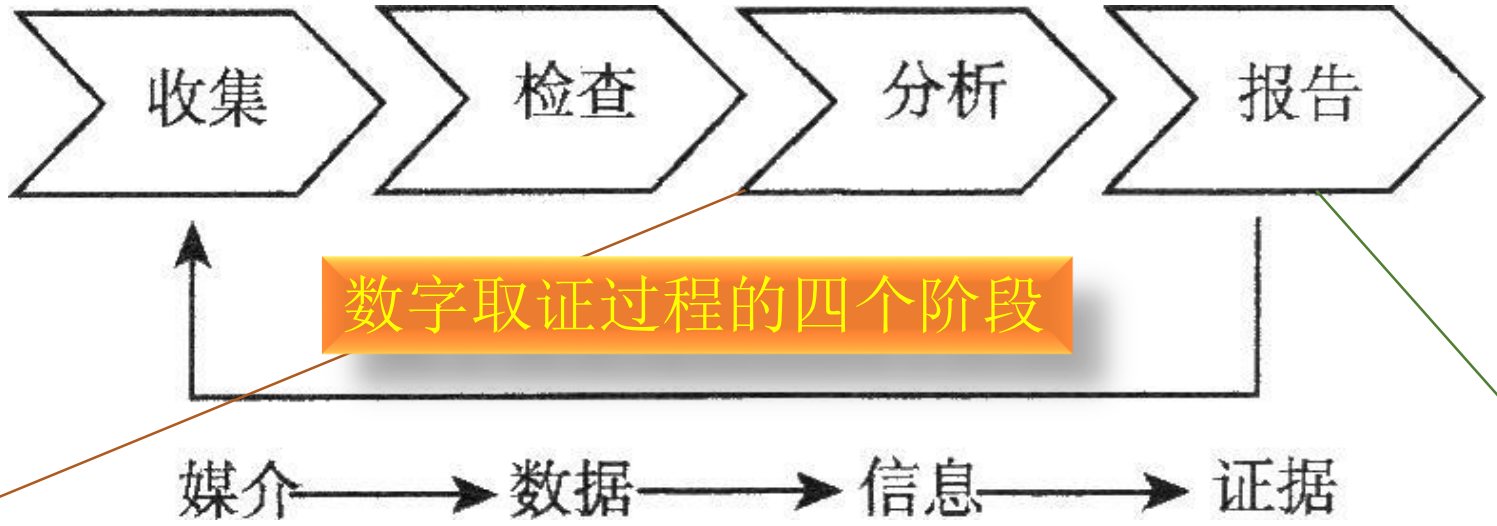


在收集阶段，就是辨认、标志、记录、集中与具体事件相关的数据，并保护数据的完整性；

在检查阶段，利用合适的取证技术和工具从收集来的数据中发现和提取相关信息，同样也要保证数据的完整性，检查阶段一般结合取证工具用手工方式进行；

数字取证过程的四个阶段

图9-3



在分析阶段，要根据特定问题和需求对检查阶段得到的数据进行进一步的调查；

最后，在报告阶段，根据分析结果，描述某个事件如何发生，决定还要采取何种行动，提出改进安全策略、安全指南、实施步骤和取证工具的建议，以及完成取证过程的其他任务。

取证过程的进一步描述

1. 收集

这个阶段的主要任务就是发现潜在的数据源并从中获取数据。

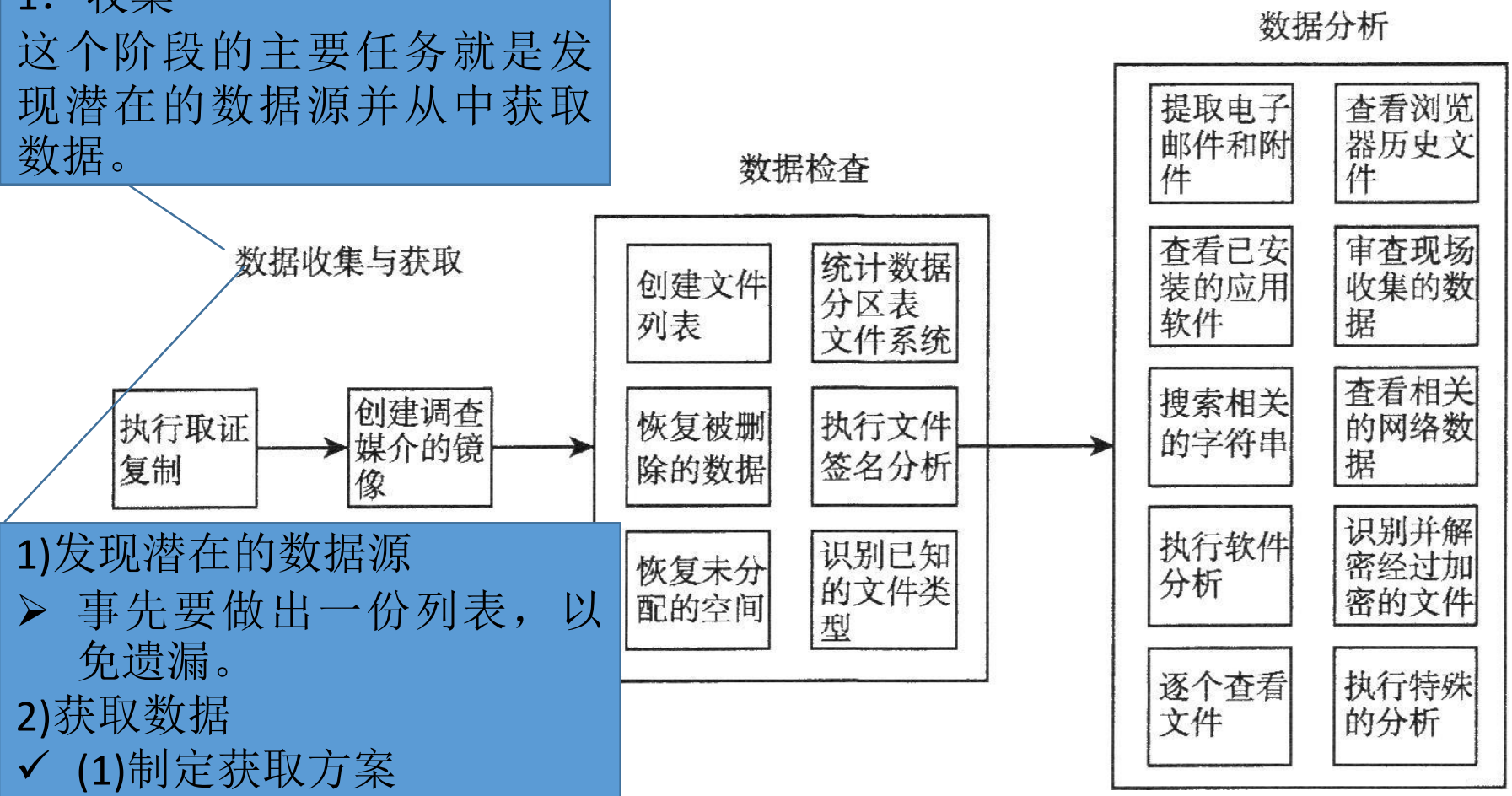


图9-4 取证的过程

1)发现潜在的数据源

- 事先要做出一份列表，以免遗漏。

2)获取数据

- ✓ (1)制定获取方案
- ✓ (2)获取数据
- ✓ (3)检验数据完整性

取证过程的进一步描述

2．检查
通过评估数据与特定事件的关联性，从收集的数据中提取信息。

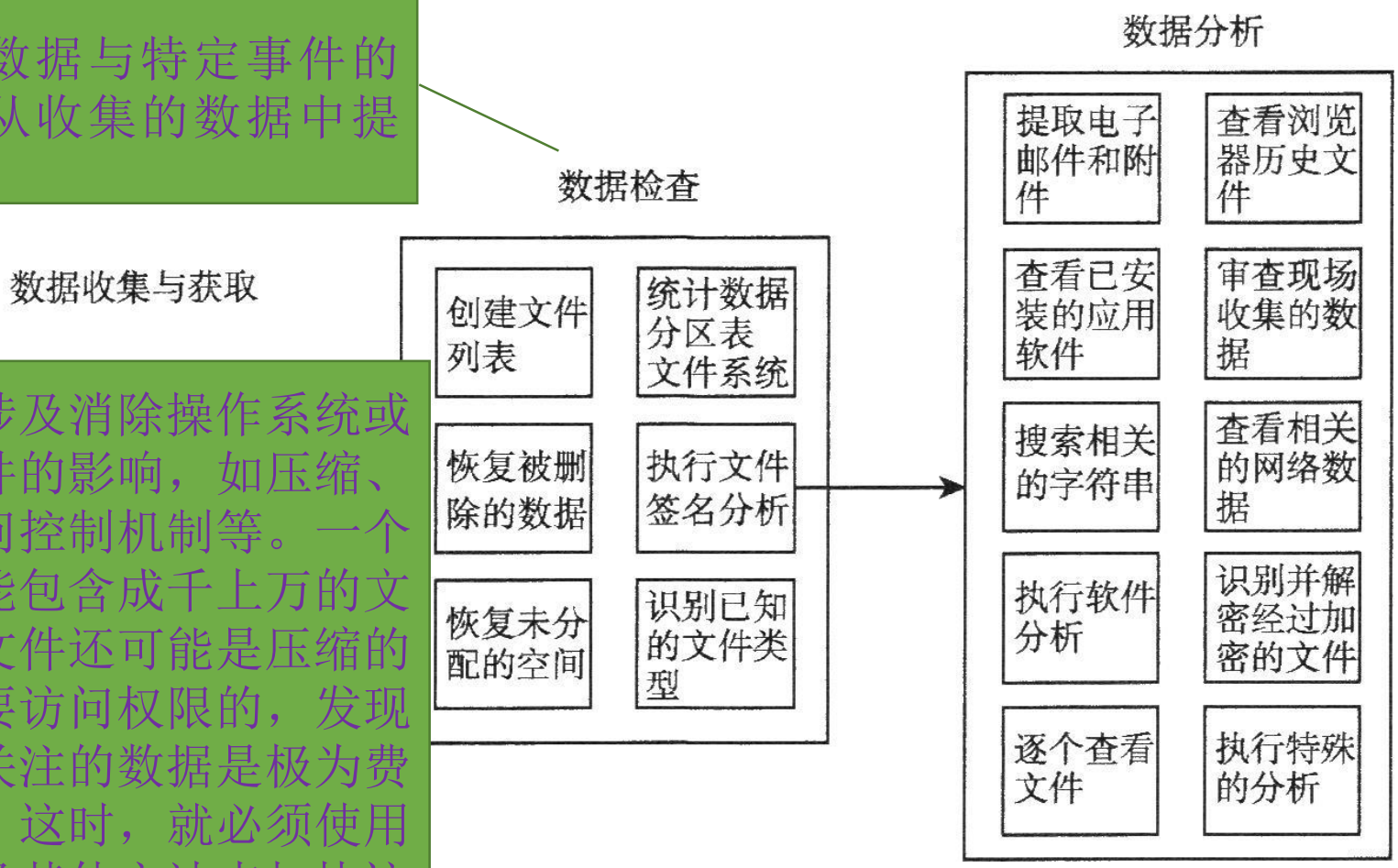


图9-4 取证的过程

这个阶段涉及消除操作系统或者应用软件的影响，如压缩、加密、访问控制机制等。一个硬盘中可能包含成千上万的文件，而且文件还可能是压缩的或者是需要访问权限的，发现其中值得关注的数据是极为费时的工作。这时，就必须使用过滤器以及其他方法来加快这个过程。

取证过程的进一步描述

3．分析

分析提取的数据进而依据系统的方法得出结论。分析工作包括辨别是谁、在什么地点、有什么物品和发生什么事件，决定这些元素如何关联并导出结论。

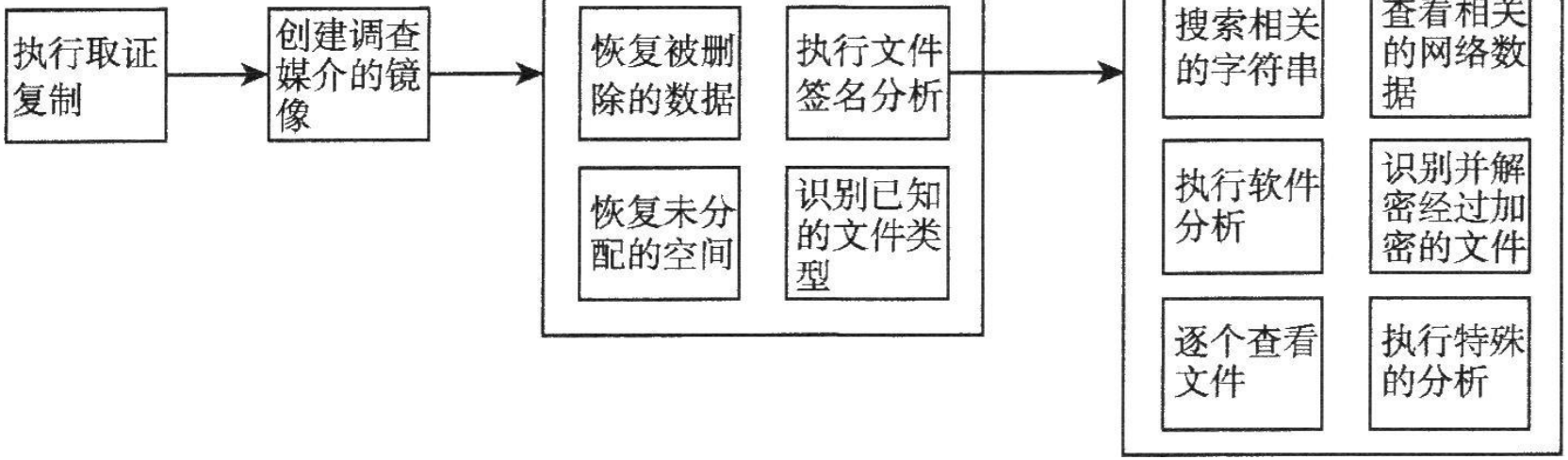


图9-4 取证的过程

取证过程的进一步描述

4．报告

- 依据分析写出报告，写报告要考虑下面的因素。
- **(1)不同的解释**：当某件事件的信息无法完整收集时，就无法确认到底发生了什么事件。当一个事件有多个合理的解释时，报告中都应给出可信的解释说明。
- **(2)不同的听众**：报告可能会提交给不同的人，对于执法人员，他们更关心信息如何收集、所有的证据是否保存；对于公司管理人员，他们更关心网络流量和相关统计数据这样的总体情况；对于公司高层，他们可能更关心到底发生了什么事件以及以后如何避免类似事件发生等这些明显的问题。

9.3.1 证据信息类别

- 计算机和网络取证中的证据信息类别可以按照其数据来源划分为：
  - 来自文件的数据
  - 来自操作系统的数据
  - 来自网络的数据
  - 以及来自应用软件的数据等。
- 这种分类不是绝对的，一般情况下，取证要综合以上各种来源的数据。

9.3.2 来自文件的数据

- 文件有很多类型，如文档文件、图像、视频、应用程序等。除了标准的文件，还有一些特殊数据，它们在取证中有时也可以起到极为重要的作用。
- **(1)删除的文件**：通常，文件被删除时，其数据没有从媒介上擦除，相反地，只是目录数据结构中指向文件存储地方的信息被标记为删除。这就意味着文件仍然存储在媒介上，但不再被操作系统列出。
- **(2)松弛空间(slack space)**：如果一个文件需要的存储空间小于分配单元（簇或块），操作系统仍然为它分配整个分配单元。例如，文件大小为7kB，分配单元大小为32kB，那么就有25kB的未使用空间。这些空间可能包含遗留的数据，如部分被删除的文件。
- **(3)空闲区**：分区上未分配的空间，可能含有以前文件数据的片断。

1．收集文件：1)从媒介复制文件

- 在收集数据时，通常要产生文件或系统的多个副本：主拷贝和工作拷贝。取证工作在工作拷贝上进行。

1)从媒介复制文件

- 有两种技术用来从媒介复制文件。
- **(1)逻辑备份(logic backup)**：从逻辑卷复制文件和目录，但不能复制删除的文件或空闲区的残余数据。
- **(2)比特流映像(bit stream imaging)**：也称为磁盘映像，逐位地复制整个原始媒介，包括空闲区和松弛区的数据。比特流映像可以是磁盘到磁盘的复制或磁盘到文件的复制。前者复制整盘的内容到另外一个磁盘，后者复制整盘的内容到另外一个文件。

完成逻辑备份和比特流映像的工具

- 有很多硬件和软件可以用来完成逻辑备份和比特流映像。
  ① 硬件如Image Masster公司的 SOLO Forensics、Logicube公司的Solitaire等。硬件通常是便携式设备，带各种媒介的接口。
  ② 软件如Linux dd、Safeback，Encase、Norton ghost、Ilook等。软件则通常带有一个引导软盘或光盘，或安装一个应用程序，然后直接或者通过网络接口、并行口传送数据。
- 要注意的是，对于一个正在运行的系统，执行整个物理设备的比特流映像是不可行的，因为这个系统的文件和内存在不断变化中，但可以执行某一个逻辑区域的比特流映像。执行逻辑备份时，要注意在复制过程中文件会变化，而且被一个进程打开的文件不容易复制。例如，复制正在运行的数据库文件、正在运行的网站文件等。

1．收集文件： 2)数据文件完整性

- 保证原始媒介的完整性，确保数据不会更改，写保护技术能实现这个目的。例如FastBlock和SCSIBlock等，如PDBlock等。
- 在执行完备份和映像操作后，需要校验副本是否是原始媒介的精确复制。计算消息摘要值(message digest)就是用来校验和确保数据完整性的方法，通常使用MD5算法和SHA-1算法，一些复制设备和软件本身内建了计算消息摘要值的功能。
- 确保完整性的步骤为：首先计算和记录原始媒介的消息摘要值，再执行备份或映像，并计算和记录副本的消息摘要值，比较这两个值来保证数据完整性得到保护；然后再次计算原始媒介的消息摘要值并和先前计算的值比较，确保备份或映像过程没有破坏原始媒介的数据。

1．收集文件： 3)文件的MAC属性

- 文件的时间属性和其他属性，如大小、名称等一样，是文件的重要属性，是取证中必须特别关注的文件属性。时间实质包括日期(date)和时间(time)，而不单指时分秒。
- 不同的操作系统，对于文件的时间属性有不同的处理方法。UNIX维护着文件的三个时间戳，也称MAC times，记录最后修改时间、最后访问时间、最后状态改变时间等信息。对于Windows来说，同样为每个文件维持着这样三个时间戳，分别是创建时间、最后写／修改时间、最后访问时间(last access time)。
- 如果要建立整个事件发生的精确的时间戳(timeline)，就必须妥善保护好时间属性。因为，并不是所有的收集工具都会保存文件的时间。映像一般会保存文件的时间，但一些工具的逻辑备份操作在复制和计算消息摘要值的同时会改变文件的最后访问时间。

1．收集文件： 4)其他技术问题A

- (1) 数据恢复。虽然可以使用诸如Final data、EasyRecovery、Disk Director Suite. Recovery等工具来恢复被删除的文件和硬盘的数据残留。但如果媒介被专用的反取证软件处理，例如，执行wiping，刻意用随机的或者持续的“0”和“1”来反复覆盖媒介，或者对媒介进行物理破坏，那么数据恢复就必须借助于更复杂和更昂贵的技术，如磁显微镜。即使这样，数据恢复的机会仍然十分渺茫。
- (2)隐含数据的修复。一些计算机系统会允许用户隐藏文件、目录和分区。这种情况下，对磁盘结构的仔细分析就非常必要，也可以借助一些软件，如Runtime的DiskExplorer等。

1．收集文件： 4)其他技术问题B

- (3) RAID阵列数据的收集。对于硬件的RAID，有两种技术收集数据：一种是使用专门的软件来重组RAID，如EnCase，这时必须记录并使用原先RAID卡的参数配置；另外一个方法就是利用取证的引导盘启动系统，直接使用原系统的RAID阵列并保证数据的完整性。对于软件的RAID，要通过查看磁盘上保存的RAID参数来重建RAID。
- (4)关于数据编码和解码。字符集是指文字的集合，对每一个文字都给予固定的内码，主要有ASCII、UniCODE、UTF8以及各种中文字符集等。数据编码则是根据一定的算法将数据转换成需要的格式，常见的数据编码有BASE64、Quoted Printable、URL encode等。了解被调查系统所使用的字符集和编码方式是极为重要的。例如在进行字符串搜索时，如果不能使用合适的字符集，就无法得到正确的结果，即使明明知道符合条件的字符串就在其中。

2．检查数据文件

- 将数据收集并保存在另外的媒介上之后，就要对数据进行检查分析。
- 这时，使用只读锁连接计算机和存放数据的媒介仍然是必需的，这样可以防止对数据不小心的修改。
- 检查技术涉及定位包含相关信息的所有文件，包括删除的文件、在松弛区和空闲区残留的文件以及隐藏的文件等，加密和口令保护使这项工作变得复杂。

1)定位文件

- 一个硬盘的存量动辄几十、几百GB，手工定位和发现包含相关信息的文件和碎片极为费时，而且必须对文件系统有透彻的了解。一些工具和技术可以帮助自动完成这个过程，例如，将空闲空间的数据转换为文件保存、恢复删除的数据、恢复回收站的数据等。
- 以Windows回收站文件分析为例。在Windows下，默认状态下使用Windows资源管理器删除文件并非真正删除，而是移动到回收站内，当清空回收站时才真正删除文件。
- 分析回收站，可以恢复回收站内的文件，得到文件内容、删除时间、原始路径等信息。Rifiuti就是一个恢复回收站数据的工具，可以恢复文件及其原先所在的目录、原先的文件名、删除时间等信息。

2)提取数据

- 要提取文件的内容，首先必须知道文件的类型。目前的文件类型极其繁多，仅已知的有明确格式定义的文件类型就有几千种，此外还有用户自己定义的文件类型。文件类型一般通过扩展名来识别，例如通过查询一些这样的数据库。然而，文件扩展名可以被删除、修改和隐藏，所以仅仅依靠扩展名来辨别文件类型是不可靠的，更精确的方法是查看文件的头部(file header)信息，文件头是一段与文件内容独立的数据。
- 另外一种方法是使用直方图了解文件类型，依靠显示文件中所有字符的ASCII码分布的百分比来判定文件类型。例如，“space”、“a”、“e”有较高分布的，可能是文本文件；几乎是相等分布的，可能是压缩文件；其他类型的，可能是加密文件，等等。

3)使用取证工具

- 取证人员必须随时准备好取证的软件或软件包，这些软件可能是多功能或单一功能的，其作用和特性明确，而且必须被实践证明，被业界广泛接受。
- 下面列出取证工具的一些功能。
- (1)文件查看：使用文件查看器代替安装原始应用程序来打开相应文件，获取其内容。如QuickView。
- (2)解压缩文件：压缩文件在分析之前就应该解压缩
- (3)目录结构的图形化显示：直观和高效地显示目录的使用情况，如Treesize。
- 4)识别已知文件：快捷地发现所需文件，排除不重要文件。
- (5)执行串搜索和模式匹配：串搜索用来从海量数据中寻找关键字或字符串，搜索可以基于某些规则及其组合。
- (6)访问元数据(metadata)
- 常用的综合取证工具包括EnCase、X-Ways Forensics、Maresware等。

3．分析文件

- 在完成数据检查后，就是对提取数据的分析，以确定事件什么时候发生、怎样发生，造成什么后果、文件何时被修改和创建、电子邮件何时发出等关键问题，并按照时间和路线重建整个事件发生、发展的时间序列和因果关系。
- 取证人员应该对取证的技术和工具所拥有的功能和产生的后果有清楚的了解，对分析的结论做出有说服力的说明。

9.3.3 来自操作系统的数据

1. 收集操作系统数据

- 对于非易失性数据的收集，类似于文件系统的收集，而易失性数据的收集，一定要在系统关机或重新启动之前进行。

1)收集易失性数据

- 常见的易失性数据包括网络连接、登录会话、内容、运行的进程、打开的文件等。是否有必要收集以及如何收集易失性数据，应该事先通过策略以及指南做出要求，以便取证人员在现场迅速地做出恰当的决定。另外，从正在运行的系统中收集易失性数据存在固有的风险。

2)收集非易失性数据

- 在易失性数据收集完成后，就可以收集下面一些非易失性的操作系统数据，如用户名、口令、日志等。

2．检查和分析操作系统数据

- 对于正在运行的系统的进程检查和分析
  - Windows： process explorer、pslis、pulist、handle、psuptime和listdll； Linux：pstree、top、ps和lsof
- 在日志文件分析方面，有UNIX/Linux中的Swatch，Windows中的psloglist和log parser。
- 在列出计划任务方面，有Windows中的at命令、schtasks、Linux中的cat cron命令等。
- 在检查非法服务和自启动进程方面，有Windows中的autoruns，Linux中的chkconfig等。
- 在查看网络连接的端口方面，有Windows中的fport、tcpview和portrptr等。

9.3.4 来自网络的数据

1. 网络流量数据源

- 网络取证的数据可以来自多种设备和应用程序，它们可以分别从网络协议的各个层捕获。

① 防火墙和路由器

② 数据包嗅探器和协议分析器

③ 入侵检测系统

④ 远程访问系统

⑤ 安全事件管理(SEM)软件

⑥ 其他来源

- 包括DHCP服务器、网络监控软件、ISP(互联网服务提供商)记录。

2．收集网络流量数据

- 通常网络流量数据分散保存在各处。但在调查取证过程中，调查人员会就某一个特定需要，使用相同的机制来收集其他额外数据。
- 网络流量数据记录在日志文件或者数据包捕获文件里，所以收集网络流量数据就如同收集日志文件或者数据包捕获文件一样简单。

3．检查和分析网络流量数据

- 当一个关注的事件被检测到，就可以评估、提取、分析网络流量数据，以决定何种事件发生、造成了何种影响。整个过程需要一系列的检查并分析多个数据源的数据，手工关联、分析这些数据，以确定可能的目的以及事件的严重程度。

1)发现相关的事件

2)检查数据源

- 同一个事件可能被多个监控设备所察觉、记载，然而，逐一地检查这些数据源既不可行，也十分低效。高效的方法是从少数几个基本的数据源开始事件调查。

3)攻击者的确认

9.3.5 来自应用软件的数据

- 所有的应用软件都包含各种形式的可执行文件或脚本，此外，还包括下面一些组件。
- 1)配置和设置：通过修改配置和设置文件，应用软件允许用户定制软件的某方面行为。
- 2)验证：用来验证试图运行应用软件的用户身份
- 3)日志：一些应用软件还记录日志，包括事件日志、审计日志、出错日志、调试日志等。
- 4)数据：所有应用软件的设计目的就是要处理数据，数据可能临时地在内存中，以及永久或临时地在文件中，这些文件类型极多。
- 5)支持文件：包括文档、图形、图标、声音等。
- 6)应用结构：指应用软件的工作模式，包括本地、客户机/服务器(Client/Server，C/S)模式、点对点(peer-to-peer，P2P)模式等。

1．收集应用软件数据

- 与应用软件相关联的数据以普通文件、易失性数据、网络流数据等形式存在。哪些数据要收集，取决于哪些应用软件是值得关注的。
- 此外，同样类型的数据，不同的应用软件可能保存在不同的地方。
- 例如Web浏览器的历史、Cache、Cookie等数据，IExplorer和Netscape就保存在不同的地方；Email的数据，不同的邮件客户端程序就以不同的文件形式保存在不同的地方。因此在收集之前，应该对这些应用软件的工作方式有必要的了解。

2．检查和分析应用软件数据

- 使用前面介绍的工具和软件，可以检查和分析应用软件的数据。但如果应用软件是特殊的，比如用户自己编写的程序，数据分析工作就会遇到困难。
- 另外，应用软件可能会具有一些安全机制来保护敏感数据，在没有授权情况下，必须借助一些专门技巧才能分析其数据，例如国内用户使用极为普遍的即时通信软件QQ等。

第9章作业

- 作业
  3. 数字取证有哪些作用？
  5. 数字取证包括哪几个主要步骤？
  6. 数字取证中，证据信息有哪些主要类别？
- 实践（自己研究，不考核）
  - 调研一款本章列出的取证工具的主要功能，如数据恢复工具Final data、EasyRecovery、Disk Director Suite、Recovery等工具，综合取证工具EnCase、X-Ways Forensics、Maresware等。