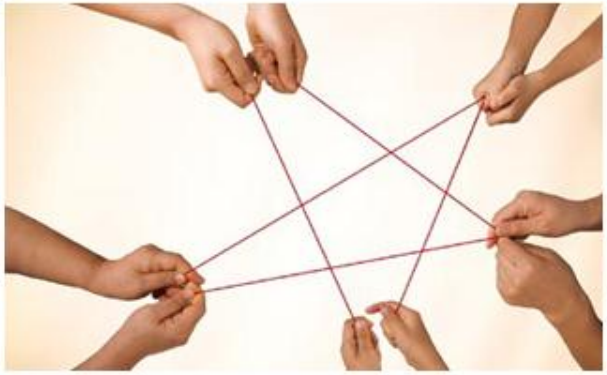
中国电信
CHINA TELECOM
世界触手可及

创新 • 激情 • 责任 • 团结



数字图书馆网络规划、安全及在线实施

中国电信集团系统集成有限责任公司

2015年7月

数字图书馆网络的定义

数字图书馆系统包括资源的采集、整理、加工和服务。

●数字图书馆资源建设是依靠传统图书馆业务网络为基础的。

●数字图书馆对外服务所依靠的网络环境又是相对独立的。

●狭义的数字图书馆的网络定义就是指数字图书馆对外服务依靠的网络环境。

数字图书馆网络的特点

●对外服务的高并发性需求

以文津搜索为例，高并发时可达到每分钟10万次。

●对外服务的快速响应需求

互联网网站访问行为统计，如果一个网站的响应速度超过3秒以上，用户对这个网站的好感度直线下降，除非用户迫切必须得到的内容，否则一般不会再理会这个网络服务。

●对外服务的高可靠性需求

当系统经常出现无响应、返回信息错误、宕机等现象，用户会认为该系统很不成熟，日后放弃使用。

●安全性需求

服务提供是面向公网上的用户，所以需要面对互联网上的DDos、TCP、UDP、DNS泛洪攻击等，所以网络安全是服务系统正常运行的保证。

数字图书馆网络规划

●高并发性、高响应性、高可靠性、安全性的网络特征与传统图书馆的业务网络特性是不一样的，因此在原有的业务网络的基础上构建数字图书馆网络环境是比较麻烦的，也是效率不高的。

●可行的方法是为数字图书馆单独构建一个网络环境，具体方法是它放在图书馆业务网的DMZ区。

●数字图书馆网络系统包括核心交换机和接入交换机，为两层结构。

●核心交换机采用双机热备方式。

●接入交换机层面：服务器采用双链路上行至接入交换机，交换机双链路上行至核心交换机。

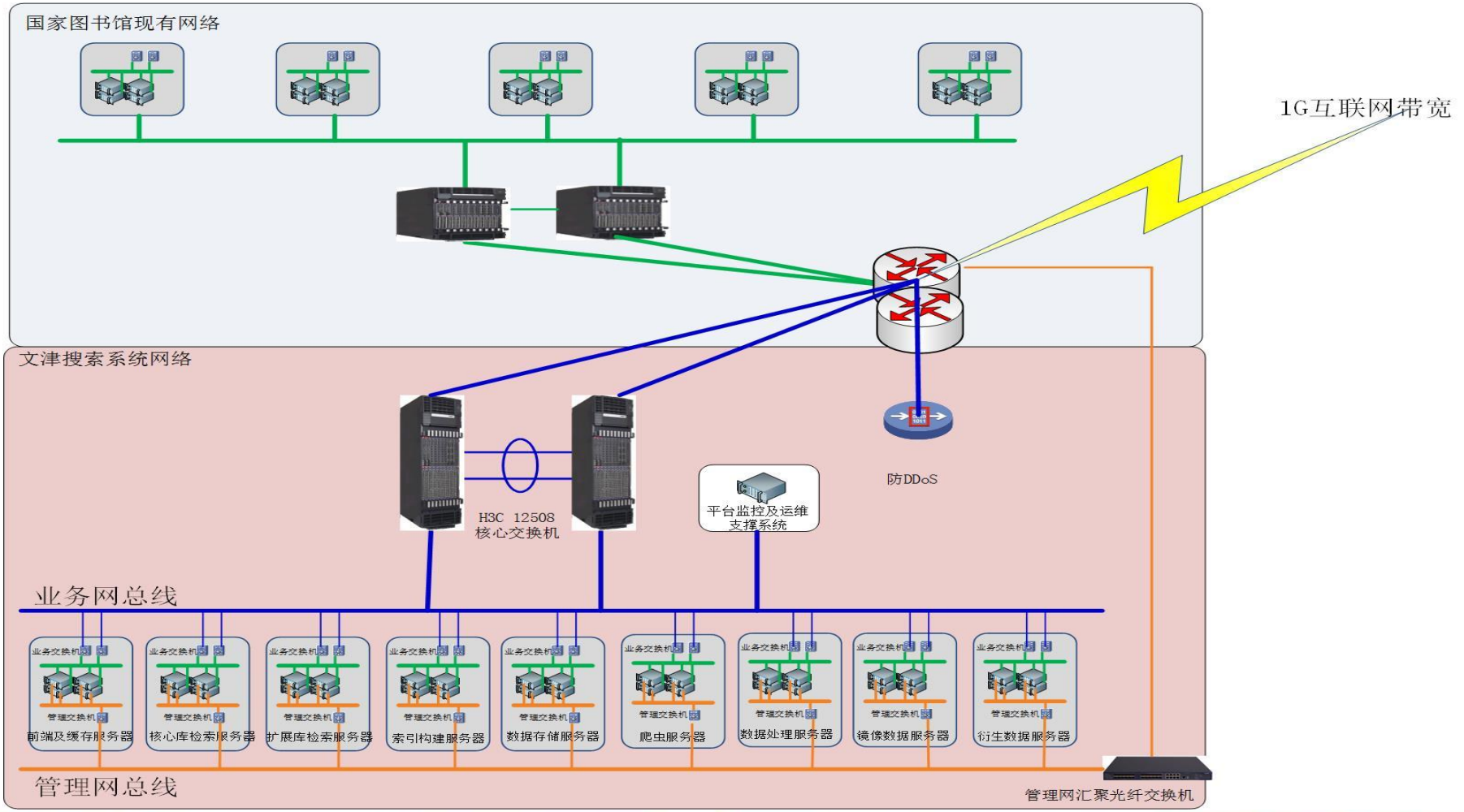
●如果条件允许就单独建立防DDos系统；如果条件不允许可借用图书馆业务网的安全系统。

实例--文津搜索网络系统

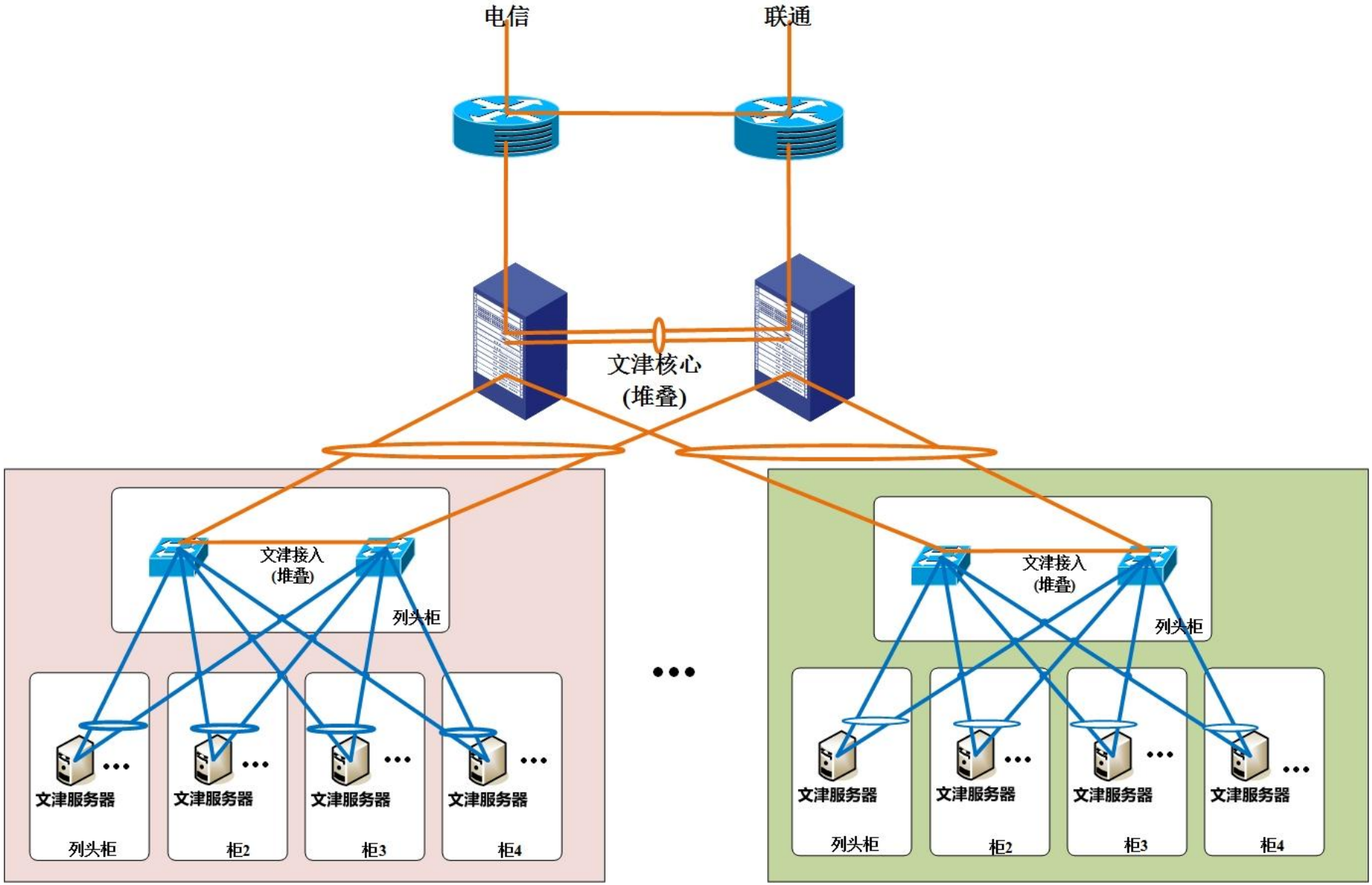
文津搜索系统是通过搜索引擎技术为国家图书馆建立一个各类元数据汇集、检索的技术支撑平台。最终建立中国图书馆届的元数据集中索引库，能够为各界提供元数据获取服务。

- ✓ 提供元数据检索，分类检索，搜索提示，拼写纠错，相关检索，全文检索等查询功能。
- ✓ 提供数据采集、分析和挖掘功能，能够提供包括互联网资源在内的用以产生搜索辅助功能所需要的数据，能基于统计分析和数据挖掘技术，满足向用户提供高质量的搜索结果排名、相关搜索和搜索建议等方面的需要。
- ✓ 具有良好的扩展性，可与其他相关网站及其它图书搜索工具连接和交互。
- ✓ 满足峰值10,000次检索请求/秒能力。在5亿条元数据，1万人并发检索请求的条件下，在1秒内满足80%的检索请求。
- ✓ 采用高可用性集群环境，动态伸缩和策略调整。随着数据量规模的增长和用户并发访问量的增加，系统可以方便的扩展以满足需求。
- ✓ 可提供1亿页全文检索能力，并可以进行古籍及其相关资源网站的全文检索。

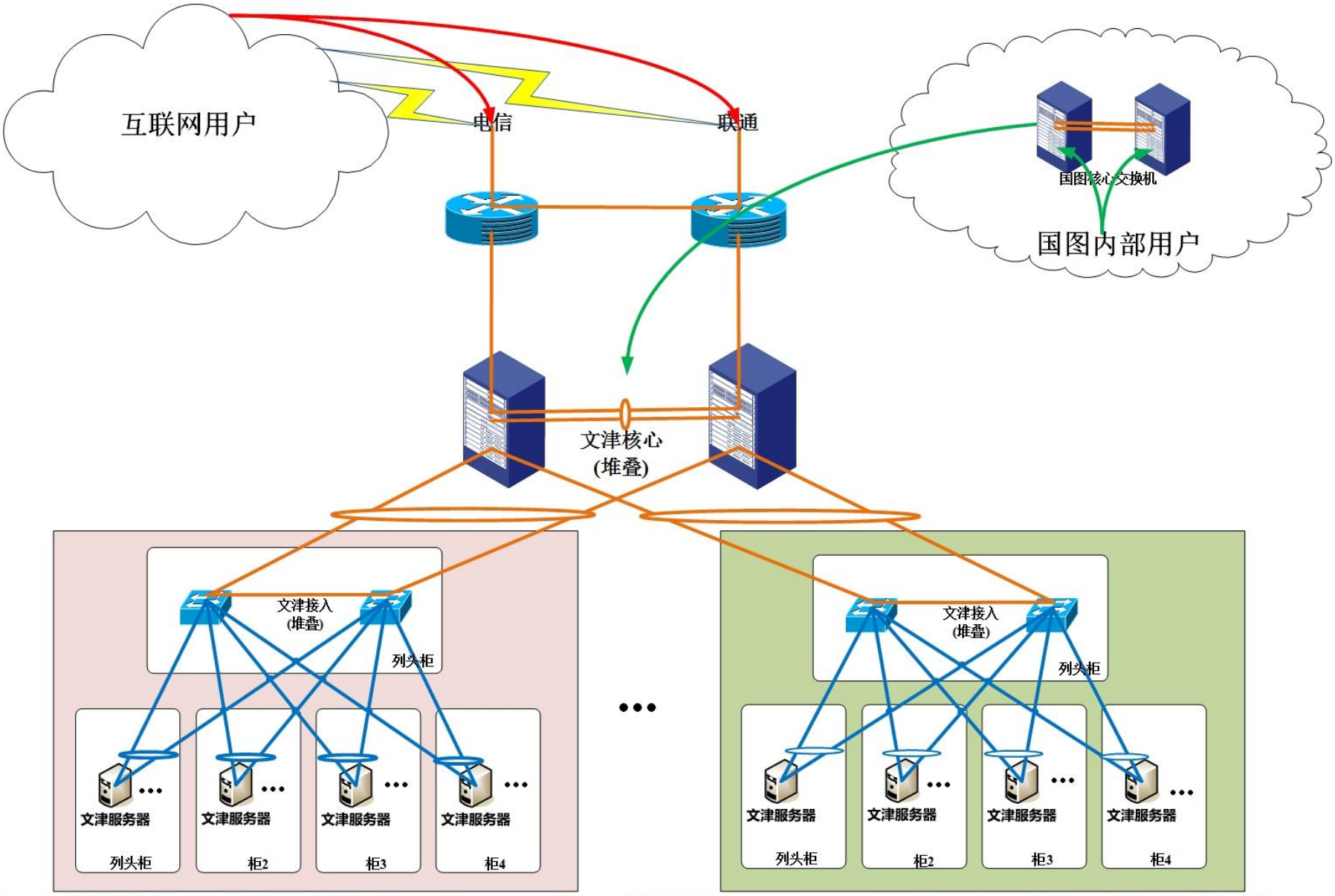
实例--文津搜索网络系统


国家图书馆现有网络
1G互联网带宽
文津搜索系统网络
H3C 12508
核心交换机
防DDoS
平台监控及运维
支撑系统
业务网总线
前端及缓存服务器
核心库检索服务器
扩展库检索服务器
索引构建服务器
数据存储服务器
爬虫服务器
数据处理服务器
镜像数据服务器
衍生数据服务器
管理网总线
管理网汇聚光纤交换机

数字图书馆规划、安全及在线实施

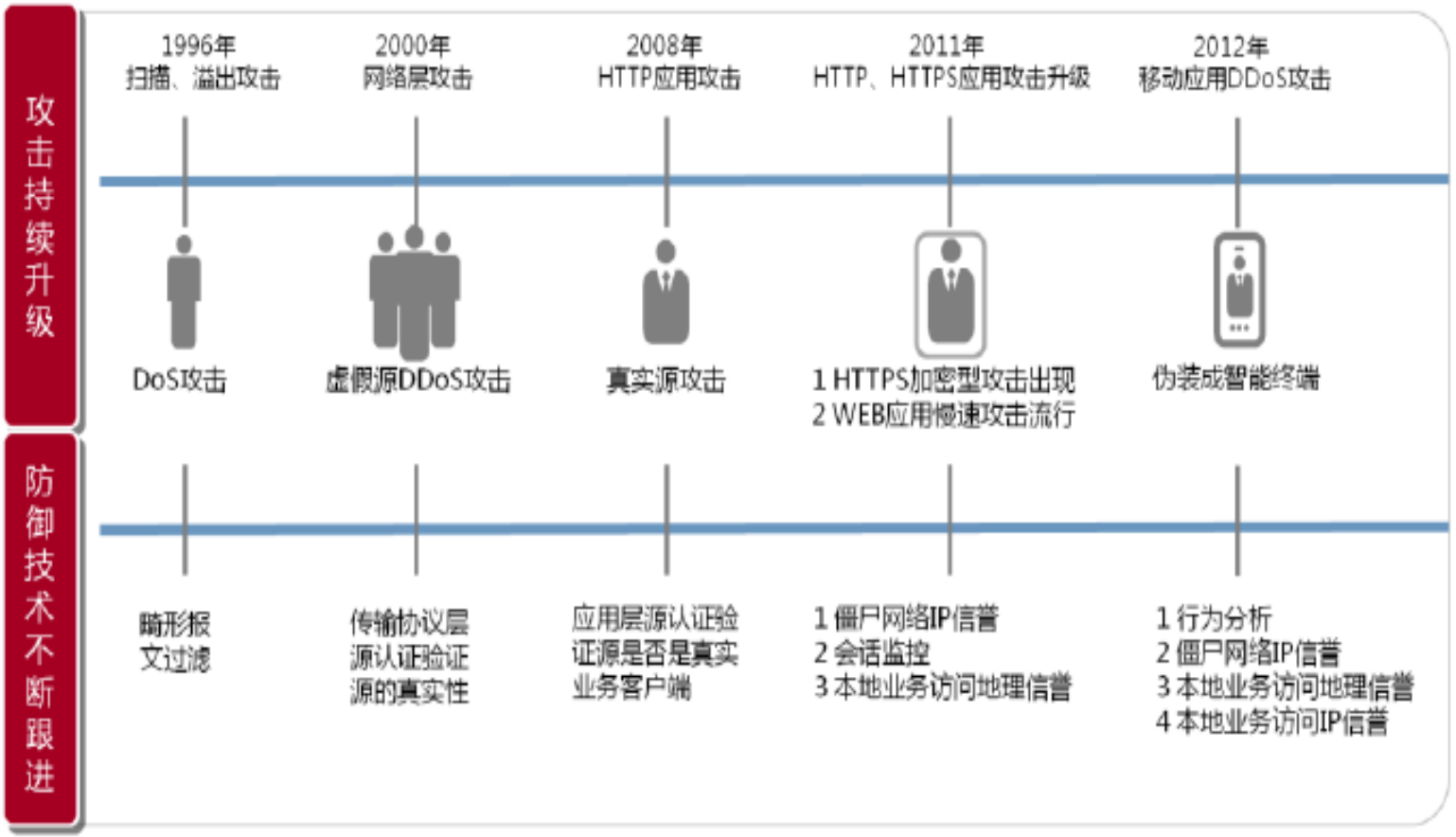


数字图书馆规划、安全及在线实施



数字图书馆网络安全

从二十世纪90年代，随着互联网的蓬勃发展，网络攻击从实验室走向了Internet，随后不断发展，攻击手段层出不穷，相应的防御攻击手段也在不断提高。



攻击和防御技术发展史

数字图书馆网络安全--DOS

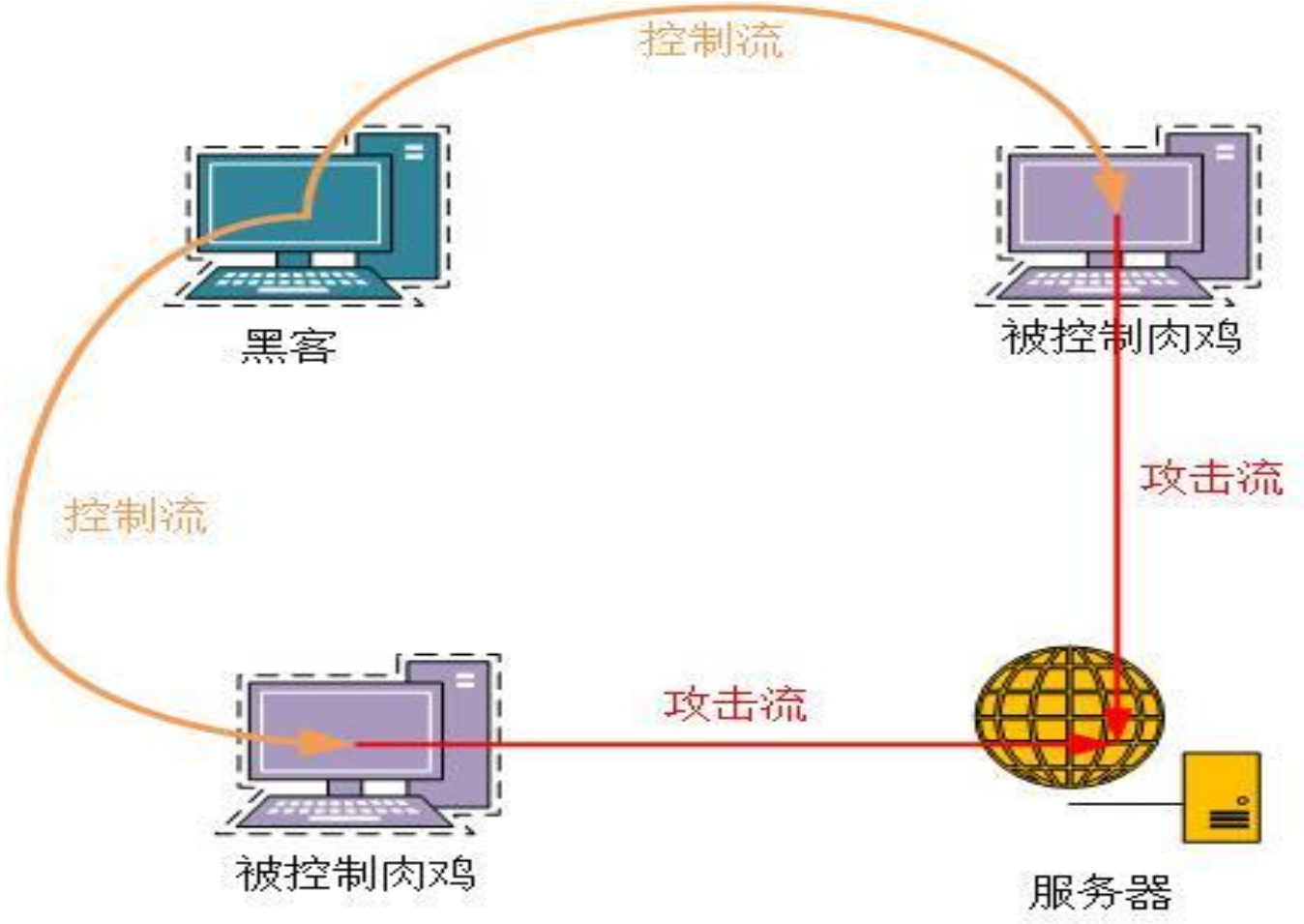
什么是DOS?

DoS 是Denial of Service 的简称，即拒绝服务，造成DoS 的攻击行为被称为DoS 攻击，其目的是使计算机或网络无法提供正常的服务。

最常见的DOS攻击就是单包攻击，一般都是以个人为单位的攻击者发动的，攻击报文比较单一。单包攻击分为以下三大类：

数字图书馆网络安全--DDOS

DDOS攻击是指攻击者通过控制大量的“肉鸡”，向被攻击目标发送大量精心构造的攻击报文，造成被攻击者所在的网络的链路拥塞、系统资源耗尽，从而使被攻击者产生拒绝向正常服务的请求提供服务的效果。



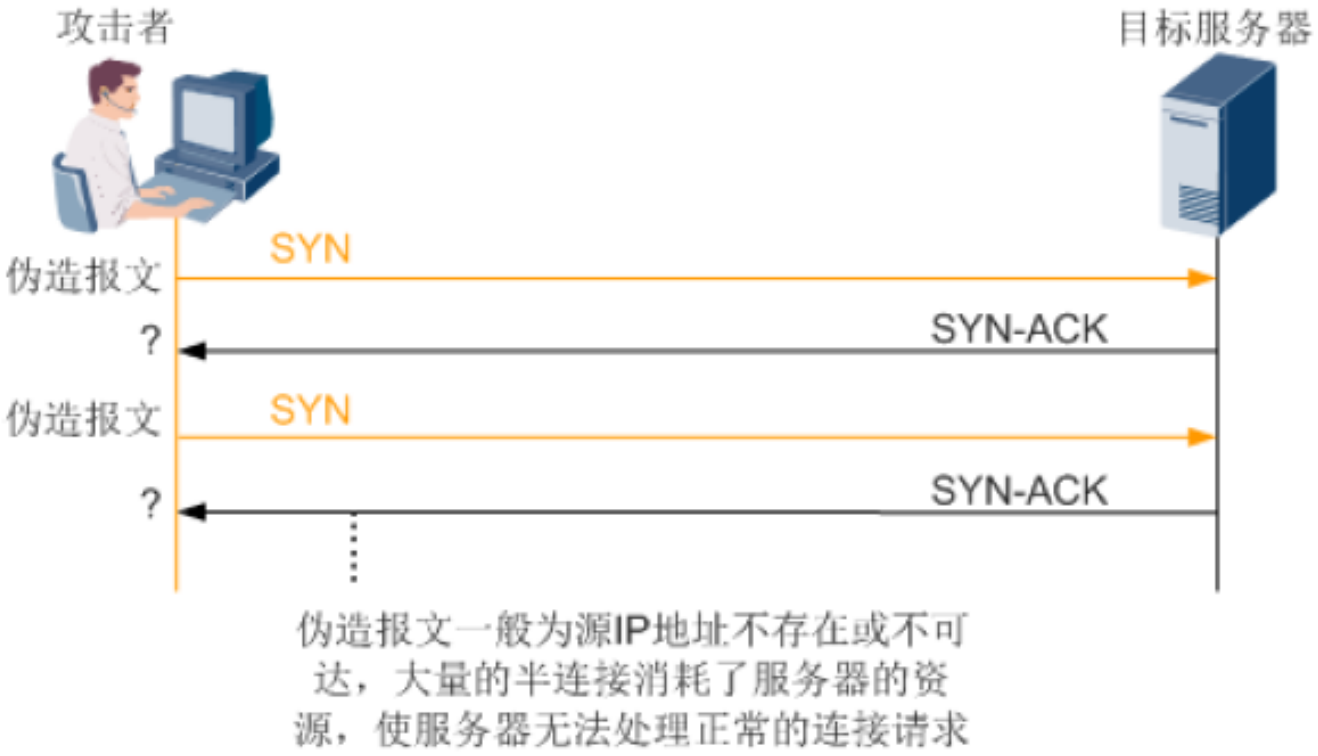
DDOS攻击是指攻击者通过控制大量的“肉鸡”，向被攻击目标发送大量精心构造的攻击报文，造成被攻击者所在的网络的链路拥塞、系统资源耗尽，从而使被攻击者产生拒绝向正常服务的请求提供服务的效果。

最初的SYN Flood 攻击类似于协议栈攻击，在当年的攻击类型中属于技术含量很高的“高大上”。当年由于系统的限制以及硬件资源性能的低下，称霸DDoS 攻击领域很久。它与别人的不同在于，你很难通过单个报文的特征或者简单的统计限流防御住它，因为它“太真实”、“太常用”。

SYN Flood 具有强大的变异能力，在攻击发展潮流中一直没有被湮没，这完全是他自身的优秀基因所决定的：

- 单个报文看起来很“真实”，没有畸形。
- 攻击成本低，很小的开销就可以发动庞大的攻击。

数字图书馆网络安全--流量型攻击之SYN Flood 及防御



攻击原理图

常用防御方法：

TCP 代理

TCP 代理是指我们的防火墙部署在客户端和服务器中间，当客户端向服务器发送的SYN 报文经过防火墙时，防火墙代替服务器与客户端建立三次握手。一般用于报文来回路径一致的场景。

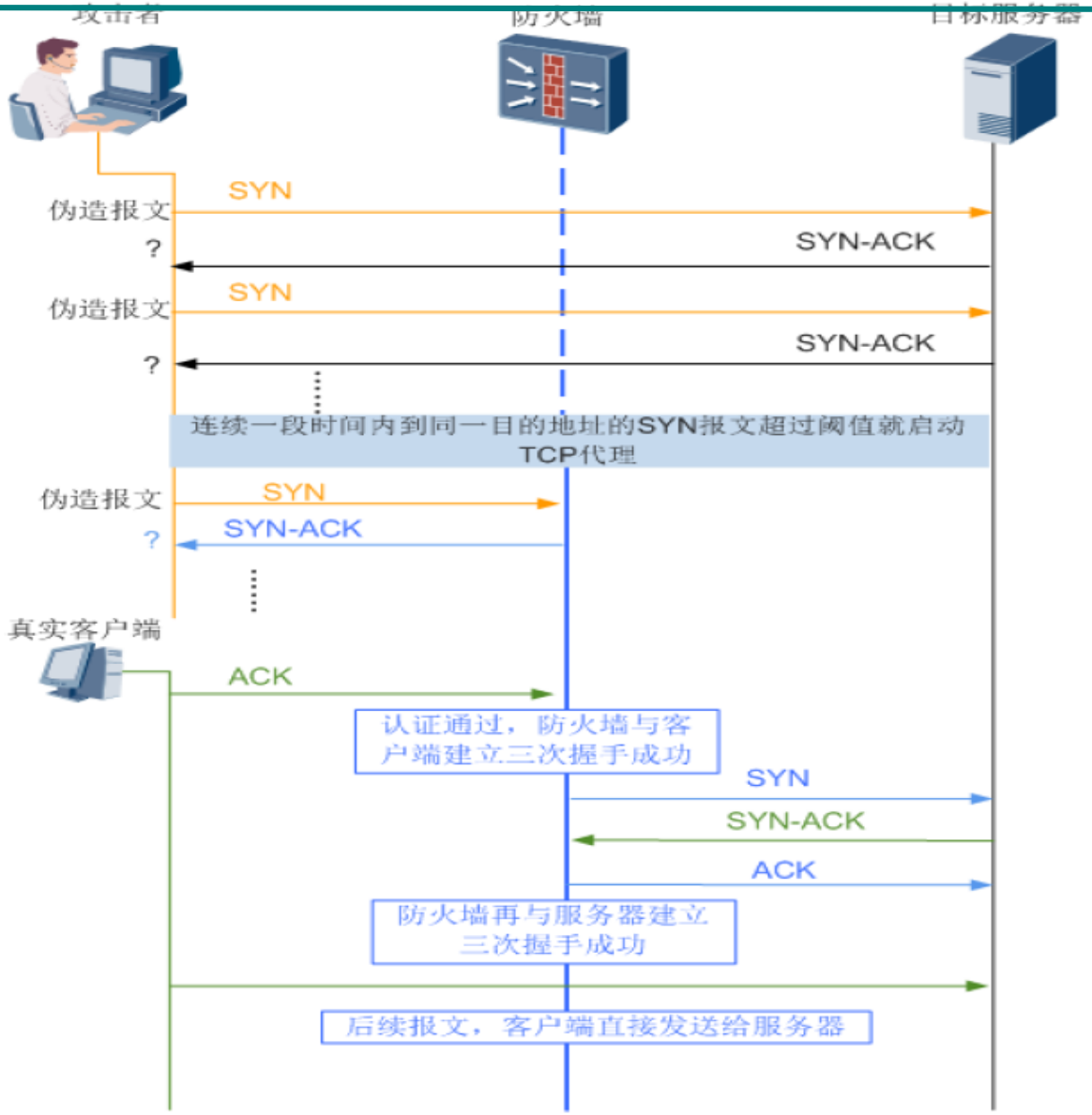
原理如下：

防火墙收到SYN 报文，对SYN 报文进行拦截，代替服务器回应SYN+ACK 报文。

如果客户端不能正常回应ACK 报文，则判定此SYN 报文为非正常报文，防火墙代替服务器保持半连接一定时间后，放弃此连接。

如果客户端正常回应ACK 报文，防火墙与客户端建立正常的三次握手，则判定此SYN报文为正常业务报文，非攻击报文。防火墙立即与服务器再建立三次握手，此连接的后续报文直接送到服务器。

数字图书馆网络安全--流量型攻击之SYN Flood 及防御



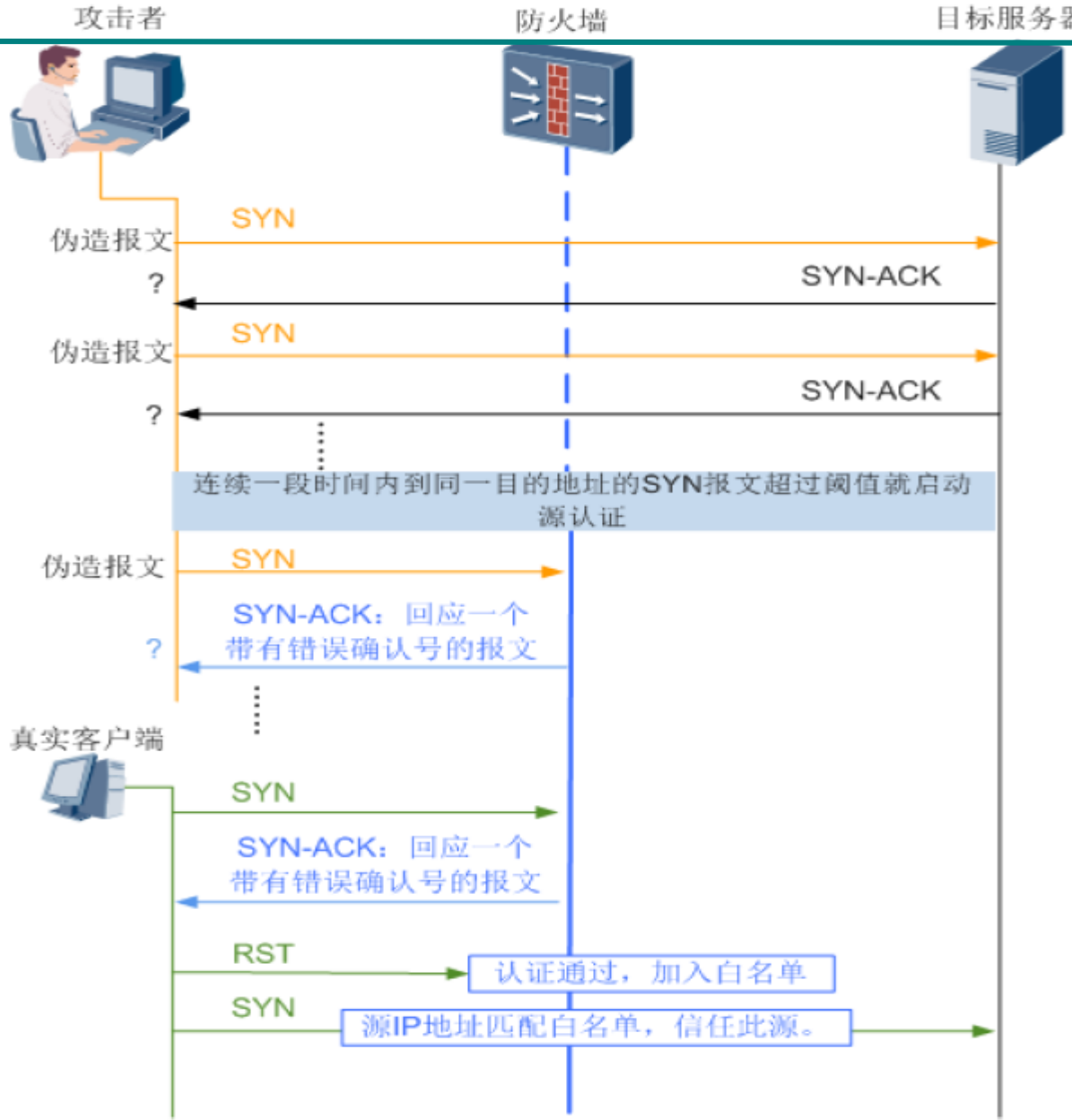
TCP 源探测

TCP 源探测是防火墙防御SYN Flood 攻击的另一种方式，没有报文来回路径必须一致的限制，所以应用普遍。

原理如下：

- 当防火墙收到客户端发送的SYN 报文时，对SYN 报文进行拦截，并伪造一个带有错误序列号的的SYN+ACK 报文回应给客户端。
- 如果客户端是虚假源，则不会对错误的SYN+ACK 报文进行回应。
- 如果客户端是真实源发送的正常请求SYN 报文，当收到错误的SYN+ACK 报文时，会再发出一个RST 报文，让防火墙重新发一个正确的SYN+ACK 报文；防火墙收到这个RST报文后，判定客户端为真实源，则将这个源加入白名单，在白名单老化前，这个源发出的报文都认为是合法的报文，防火墙直接放行，不在做验证。

数字图书馆网络安全--流量型攻击之SYN Flood 及防御



UDP Flood 属于带宽类攻击，黑客们通过僵尸网络向目标服务器发起大量的UDP 报文，这种UDP 报文通常为大包，且速率非常快，通常会造成以下危害：

- 消耗网络带宽资源，严重时造成链路拥塞。
- 大量变源变端口的UDP Flood 会导致依靠会话转发的网络设备，性能降低甚至会话耗尽，从而导致网络瘫痪。

防火墙对UDP Flood 的防御并不能像SYN Flood 一样，进行源探测，因为它不建立连接。常见的防御方法有两种：

限流

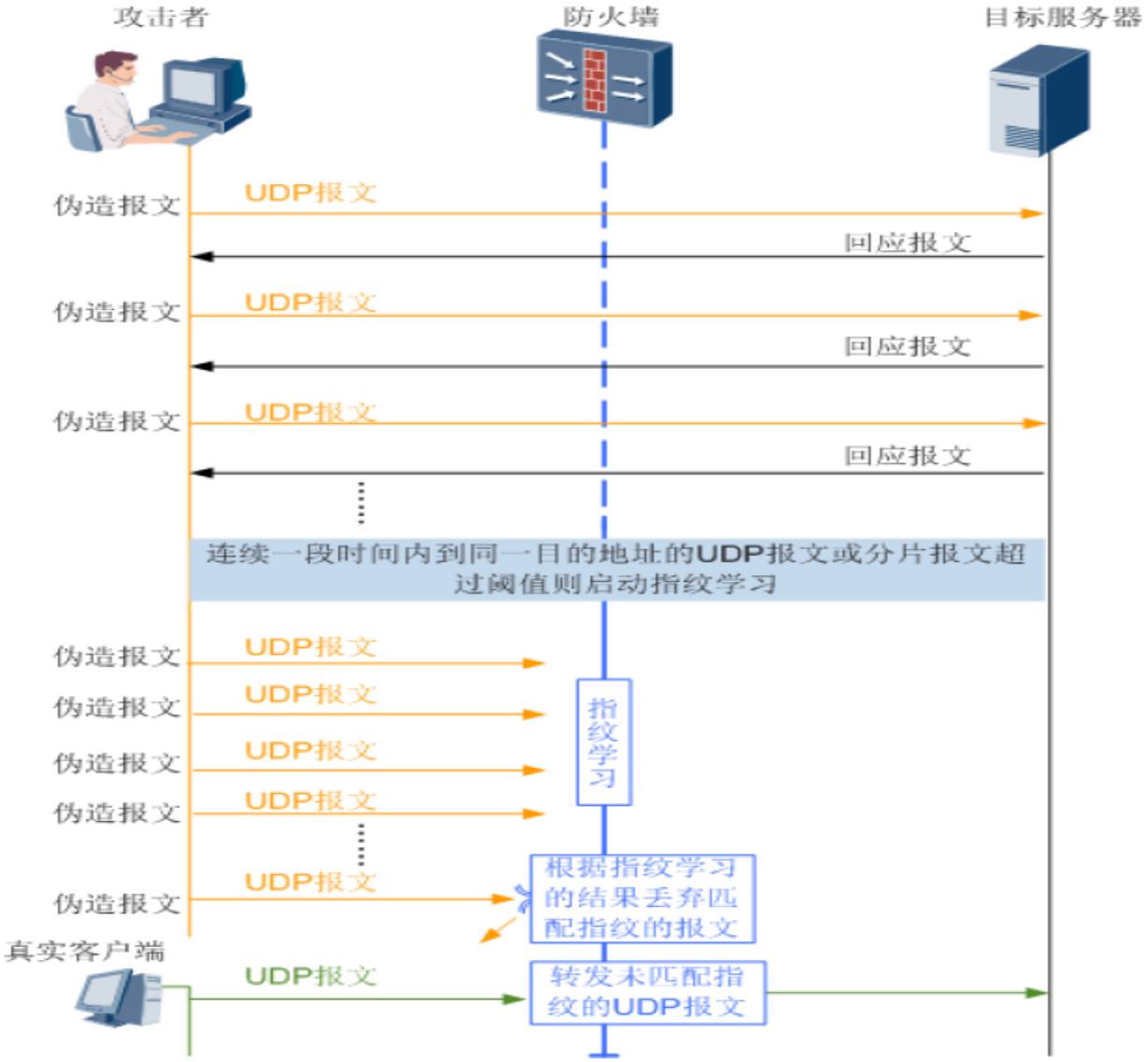
●基于流量入口的限流：以某个接口流量作为统计对象，对通过这个接口的流量进行统计并限流，超出的流量将丢弃。

●基于目的IP 地址的限流：即以某个IP 地址作为统计对象，对到达这个IP 地址的UDP流量进行统计并限流，超过部分丢弃。

●基于目的安全区域的限流：即以某个安全区域作为统计对象，对到达这个安全区域的UDP流量进行统计并限流，超过部分丢弃。

●基于会话的限流：即对每条UDP 会话上的报文速率进行统计，如果会话上的UDP 报文速率达到了告警阈值，这条会话就会被锁定，后续命中这条会话的UDP 报文都被丢弃。当这条会话连续3 秒或者3 秒以上没有流量时，防火墙会解锁此会话，后续命中此会话的报文可以继续通过。

指纹学习


攻击者
防火墙
目标服务器
伪造报文
UDP报文
回应报文
伪造报文
UDP报文
回应报文
伪造报文
UDP报文
回应报文
连续一段时间内到同一目的地址的UDP报文或分片报文超过阈值则启动指纹学习
伪造报文
UDP报文
伪造报文
UDP报文
伪造报文
UDP报文
伪造报文
UDP报文
指纹学习
伪造报文
UDP报文
根据指纹学习的结果丢弃匹配指纹的报文
真实客户端
UDP报文
转发未匹配指纹的UDP报文

常见的DNS Flood 攻击一般都是攻击者向DNS 服务器发送大量的不存在的域名解析请求，导致DNS 缓存服务器不停向授权服务器发送解析请求，最终导致DNS 缓存服务器瘫痪，影响对正常请求的回应。

DNS 服务器支持TCP 和UDP 两种协议的查询方式，但是大数的查询都是使用UDP 查询的，当发生DNS Flood 攻击时，防火墙收到DNS 请求，会代替DNS 服务器响应DNS，要求DNS 客户端以TCP 方式发送DNS 请求。

- 如果客户端是真实源，会继续以TCP 方式发送DNS 请求。
- 如果客户端是虚假源，则不会再以TCP 方式发送DNS 请求。

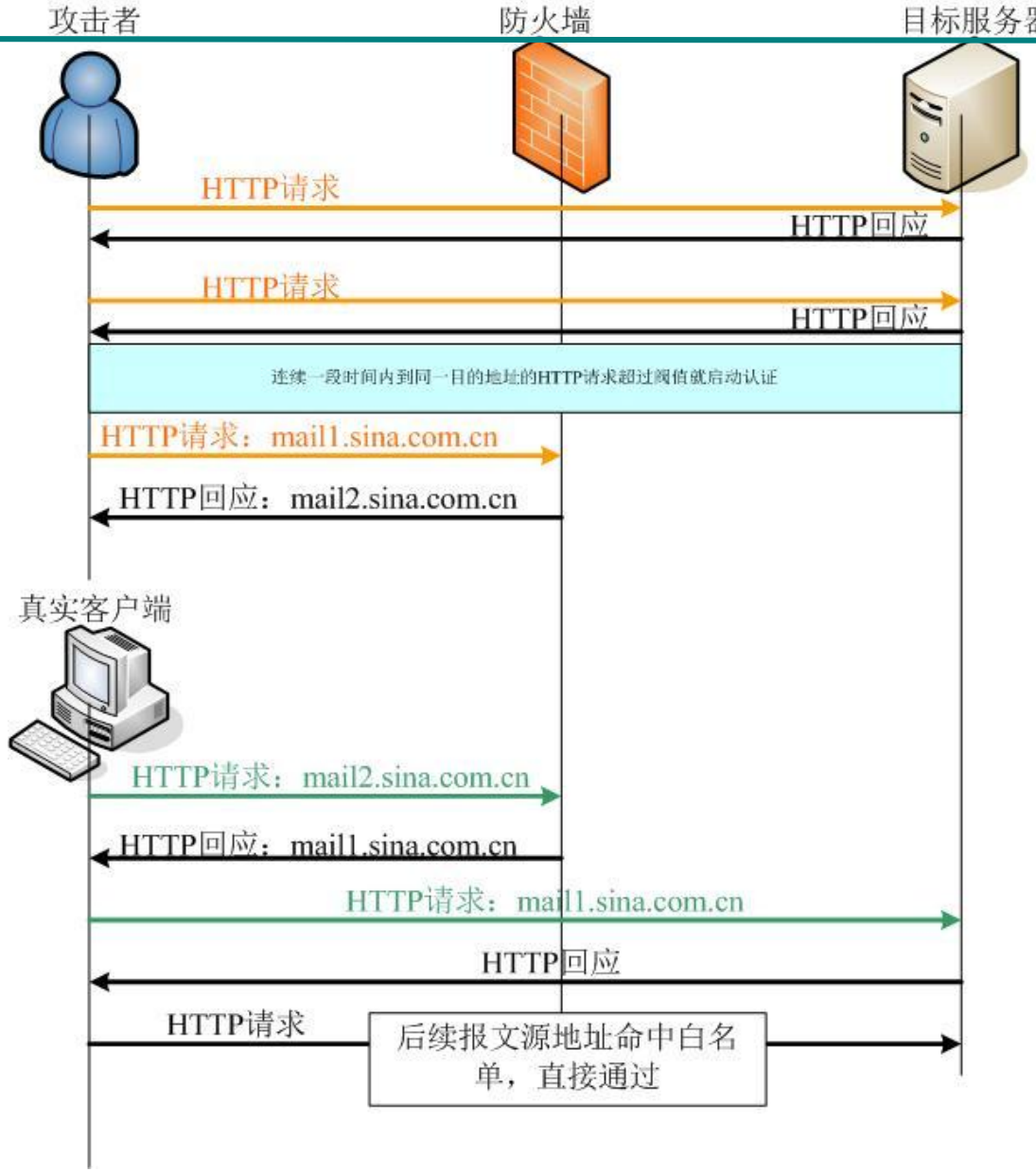
常见的HTTP Flood 攻击，一般指黑客通过代理或僵尸主机向目标服务器发起大量的HTTP报文，这些请求涉及数据库操作的URI 或其它消耗系统资源的URI，目的是为了造成服务器资源耗尽，无法响应正常请求。

防火墙对于HTTP Flood 的防御，主要依靠HTTP 协议所支持的重定向方式，譬如说客户端向服务器请求www.sina.com，服务器可以返回一个命令，让客户端改为访问www.sohu.com。

这种重定向的命令在HTTP 协议栈中是合法的。我们防火墙的防御机制就是利用这个技术点，来探测HTTP 客户端是否为真实存在的主机。

具体的防御过程如下：

数字图书馆网络安全--应用层攻击（HTTP Flood）及防范



实例--文津搜索网络系统

1、限制网络端口；

关闭所有不必要的网络端口，仅仅开放对外服务端口、特定的维护端口；

2、定制操作系统；

通过对操作系统中可能存在安全隐患部分进行定制修改，优化系统内核，使系统更加安全、高效；

3、通过访问控制策略严控可暴露主机；

通过IPTables进行访问策略控制，只允许外网访问内网的特定服务器。

4、ACL控制

根据实际运行环境，在路由器上配置相应的ACL。

数字图书馆规划、安全及在线实施

在线实施：

数字图书馆的网络建设是一定不能影响现有的图书馆的网络的正常运行，这就涉及到一个在线施工的问题。

中国电信承担了国家信息中心、国家安管中心、财政部、公安部、国安委、国家保密局、国家税务总局、环保部、中科院等单位的网络建设和改造，积累了丰富的在线实施经验。国家图书馆2008年的二期网络和2014年的南网改造也是由中国电信承担的，所以在这里跟大家分享在线实施的经验。

数字图书馆网络系统的在线实施

1、首先根据网络规划搭建一套独立的网络运行环境。

2、与现有的业务网络并网，并网是整个项目的关键点。

3、从下往上逐步完成。即首先完成接入交换机的替换和升级，最后完成核心交换机的升级和替换。核心交换机的割接是整个项目的关键，在这之前需要做多次的模拟测试。

4、割接时日工作量确定。以当天下班时间到第二天上班时间为一个基本工作日，按工作日制定实施计划。为做到计划准确，要提前计算出每个工作日工作量（包括交换机的更换、跳线、打签等），在白天非工作日时间完成交换机的配置设置和测试。为此需要建立一个调试实验区，在这方面，中国电信有巨大的优势，有冗余的设备和环境供模拟测试。

5、制定详细的回退方案。如当上班前3小时，发现更换和割接工作无法进行，就必须回退到工作前的状态。

谢 谢！

Thanks