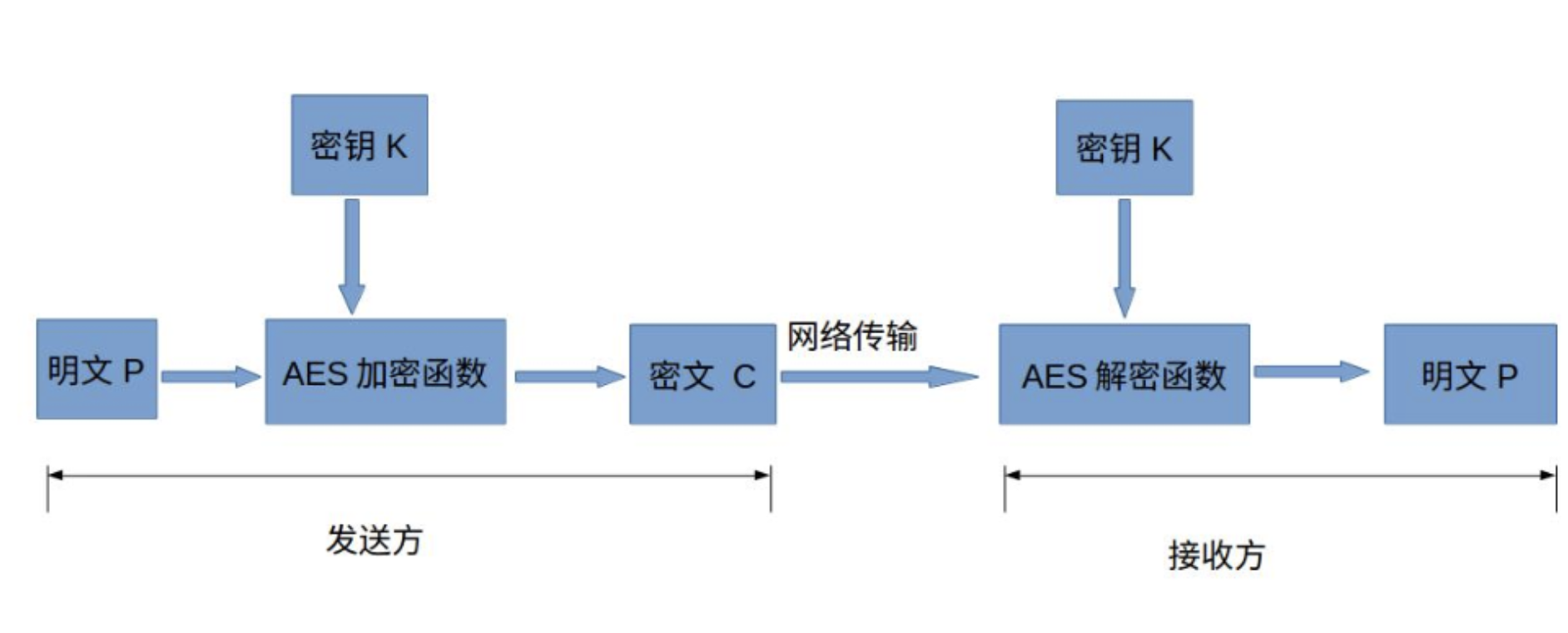
# 爬虫面试笔记

# 1 加密算法：

AES对称加密：



<https://blog.csdn.net/qq_28205153/article/details/55798628>

RSA非对称加密：

非对称加密和对称加盟店区别在于加密时密钥为公钥，解密时为私钥

2三次握手四次挥手

客户端向服务器发送消息报文请求连接

服务器收到请求后，回复报文确定可以连接

客户端收到回复，发送最终报文连接建立

主动方发送报文请求断开连接

被动方收到请求后，立即回复，表示准备断开

被动方准备就绪，再次发送报文表示可以断开

主动方收到确定，发送最终报文完成断开

3 程是操作系统分配资源的最小单元, 线程是操作系统调度的最小单元。

一个应用程序至少包括1个进程，而1个进程包括1个或多个线程，线程的尺度更小。

每个进程在执行过程中拥有独立的内存单元，而一个线程的多个线程在执行过程中共享内存。

# 4 在网页里输入域名，按下确定的过程后，发生了什么

1.域名解析

域名解析检查顺序为：浏览器自身DNS缓存---》OS自身的DNS缓存--》读取host文件--》本地域名服务器--》权限域名服务器--》根域名服务器。如果有且没有过期，则结束本次域名解析。域名解析成功之后，进行后续操作

2.tcp3次握手建立连接

3.建立连接后，发起http请求

4.服务器端响应http请求，浏览器得到到http请求的内容；

5.浏览器解析html代码，并请求html代码中的资源

1. 浏览器对页面进行渲染，展现在用户面前

# 5 Cookie的机制

在网站中，http请求是无状态的，也就是说，即使第一次和服务器连接后并且登录成功后，第二次请求服务器依然不能知道当前请求是哪个用户。cookie的出现就是为了解决这个问题：当浏览器访问网站后，这些网站将一组数据存放在客户端，当该用户发送第二次请求的时候，就会自动的把上次请求存储的cookie数据自动携带给服务器，服务器通过浏览器携带的数据就能识别当前用户。

cookie中的Domain和Path属性标识了这个Cookie是哪一个网站发送给浏览器的；Expires属性标识了cookie的有效时间：

1）如果对Expires属性进行有效时间设置，当cookie的有效时间过了之后，这些数据就被自动删除了；此时会话cookie保存在硬盘上。关闭浏览器后再次打开，这些cookie依然有效，直到超过设定的有效时间。存储在硬盘上的cookie可以在不同的浏览器进程间共享，比如两个IE窗 口。

2）如果不设置过期时间，则表示这个cookie生命周期为浏览器会话期间，只要关闭浏览器窗口，cookie就消失。这种生命期为浏览会话期的cookie被称为会话cookie。会话cookie一般不保存在硬盘上，而是保存在内存里。

特点：cookie存储在本地浏览器，且存储数据量有限，不同的浏览器有不同的存储大小，但一般不超过4KB。因此使用cookie只能存储一些小量的数据。

Session的机制

Session是存放在服务器端的类似于HashTable结构（每一种Web开发技术的实现可能不一样，下文直接称之为HashTable）来存放用户 数据，当浏览器第一次发送请求时，服务器自动生成了一个HashTable和一个Session ID用来唯一标识这个HashTable，并将其通过响应发送到浏览器。当浏览器第二次发送请求，会将前一次服务器响应中的Session ID放在请求中一并发送到服务器上，服务器从请求中提取出Session ID，并和保存的所有Session ID进行对比，找到这个用户对应的HashTable。

一般情况下，服务器会在一定时间内（默认20分钟）保存这个HashTable，过了时间限制，就会销毁这个HashTable。在销毁之前，程序员可以 将用户的一些数据以Key和Value的形式暂时存放在这个HashTable中。当然，也有使用数据库将这个HashTable序列化后保存起来的，这 样的好处是没了时间的限制，坏处是随着时间的增加，这个数据库会急速膨胀，特别是访问量增加的时候。一般还是采取前一种方式，以减轻服务器压力。

总而言之，session和cookie的作用是类似的，都是为了存储用户相关的信息。

特点：cookie存储在本地浏览器，而session存储在服务器。存储在服务器的数据会更加的安全，不容易被窃取。但会占用服务器的资源。

# rabbitmq消息队列

http://blog.itpub.net/69902581/viewspace-2673724/

# elasticsearch

自己找面试题

# 倒排索引

在搜索引擎中，每个文档都有一个对应的文档 ID，文档内容被表示为一系列关键词的集合。例如，文档 1 经过分词，提取了 20 个关键词，每个关键词都会记录它在文档中出现的次数和出现位置。

那么，倒排索引就是关键词到文档 ID 的映射，每个关键词都对应着一系列的文件，这些文件中都出现了关键词

# 安卓逆向方面：

## xopsed框架底层原理，脱壳工具

## openMemory的脱壳原理

# 思想，此方法是根据openMemory的地址来计算出此hook app运行时dex文件的地址，dex文件的地址加上32个字节，就是dex文件的大小

# 根据dex文件的大小和地址 就可以得到dex文件

## 安卓双向验证打印证书密码的原理

jadx反编译高度混淆找不到参数 使用objection内存漫游

## nps安卓群控

scrpy底层框架原理，改写srpay

redis集合。mysql数据库

## GIL

在 Python 程序的工作示例。其中，Thread 1、2、3 轮流执行，每一个线程在开始执行时，都会锁住 GIL，以阻止别的线程执行；同样的，每一个线程执行完一段后，会释放 GIL，以允许别的线程开始

GIL 的功能是：在 CPython 解释器中执行的每一个 Python 线程，都会先锁住自己，以阻止别的线程执行。