信息隐写--1998年出版高被引论文--on the limits of steganography隐写技术的局限性

原创

讒訟_ 于 2021-07-04 20:38:02 发布 108 收藏 2
分类专栏： 信息隐写 文章标签： 信息隐写 信息安全

本文链接：https://blog.csdn.net/weixin_41963310/article/details/118467004

这篇文章中出现最多的就是“囚犯问题”这个例子，在阅读论文中，结合这个例子及其变形来理解，可以更容易理解作者想要表达的意思。全文没有一张图，一共8页。。。看了好几天才看完

introduce

与隐藏信息的传统加密相比，隐写技术的目的是隐藏信息的存在。隐写技术只有在存在竞争对手或攻击者时才有意义，这个攻击者可能是被动攻击、只影响通信过程的，或是一个主动进行攻击、并试图篡改内容的攻击者。

一个比较有意思的例子是1586年苏格兰的玛丽女皇意图刺杀英国女皇伊丽莎白。但是玛丽女皇的密信被破解了，英国的安全警察根据被破解的密信，在玛丽女皇给反叛者头领的密信中添加了内容，从而获悉了刺杀者的名单，破坏了这一次刺杀行动。

关于隐写技术在科学领域的研究可以追溯到1983年。Simmons提出了一个"囚犯问题"：爱丽丝和鲍勃是身处监狱的囚犯，他们想要商量一个逃狱计划，但是他们之间的信件都要经过监狱长的审查，一旦被监狱长发现信件中包含隐藏信息，计划就会败露，所以他们必须要想到一个将密文隐藏在看似无害的文本中，其中的安全性完全取决于爱丽丝和鲍勃以某种方式共享的密钥的安全性。

隐写技术不应该和加密技术相混淆。加密技术是对信息进行转换，使得第三者在截取到加密信息后不知所措，但这种保护往往是不够的，比如在战场中发现士兵和敌对方进行通信（这时，只需要知道通信双方就足够了，而加密信息则没有那么重要），所以仅仅对信息进行加密是不够的。

但是身份验证也并不是万能的，因为可能是冒名顶替。在过往的一些案例中，有毒犯分接邻居家的电话以躲避监听，甚至是因为GSM（全球移动通信系统）的强大功能，犯罪人群使用盗抢的手机进行通信。

隐藏原始信息的技术包括：

1. 隐藏信息的存在；
2. 隐藏信息的有效时间；
3. 隐藏信息的发送方和接收方；

上述技术都属于流量安全技术领域，隐写技术只是其中的一个方面。

what is steganography什么是隐写

传统的隐写技术是用来在cover field（可以是图片、文字、视频、音频甚至是计算机编码等）中嵌入隐藏信息（这里的隐藏信息可能是版权信息、密码通信甚至是一串数字等）的技术。

在嵌入过程中，最关键的是密钥。在不知道密钥的情况下，第三方无法检测和改变嵌入的信息。一旦cover field被嵌入信息，则其可以被称为stego field（隐写文件）。

最近几年，隐写技术开始被广泛关注，主要原因有两个：

1. 出版社和广播公式为了在其产品中隐藏版权标志和编码开始对隐写技术感兴趣，因为数字媒体太容易被copy；
2. 很多政府采取措施来限制加密服务的可用性，促使人们研究如何将私人信息嵌入看似无害的掩盖信息中，这可以看作是对政策的一种反制。

在数字媒体中隐藏版权信息和经典的囚犯问题还有一些不同：

1. 在囚犯问题在，只要攻击者（监狱长）发现其中的包含标记即为成功；
2. 而对于版权信息而言，每个人都知道其中隐藏着版权信息，但是此时的关键是如何使其不起作用或如何对隐藏的信息进行篡改、删除。

state of the art

密码学的原则是：假定加密方法以被敌人获悉，其安全性完全依赖于密钥的选择。（Kerckhoffs 1883年提出）

simple system

很多的方法都可以在图像的最低有效位嵌入隐藏信息。信息被这样隐藏人眼是很难观察到的，但是对于第三方来说，是很容易被刻意的检测出来并篡改的。

效果稍微好一些的方法是根据发送方和接收方共享的一个密钥，使用传统的密钥流生成器将密钥扩充成一个伪随机密钥流，这个密钥流被用来选择像素或声音位置来嵌入密文信息。

并不是所有的像素都可以被用来嵌入隐藏信息的，在一些颜色变化小，或位于边界的位置，一旦嵌入信息是很容易被发现的，所以有一些系统事先检验像素周围的亮度变化，只有像素符合要求才被用来嵌入信息。

但是上述的方法隐藏的信息有很多方法进行破坏。比如大多数的滤波操作都会更改最低有效位的信息。

一种比较有效的应对措施是使用冗余——使用错误校验码（检测信息是否被破坏）或简单的嵌入大量的标记（比如嵌入大量的水印信息，即便一些遭到破坏，总有一些是完整的）；比如“Patchwork”方法通过在大量伪随机数选择的像素块中增加亮度方差来隐藏信息，上述方法同样可以应用在音频中。

对于这类方法的攻击方法也很简单，图像可以被裁剪，对于音频的话，采用的方法是随机删除一段音频，然后使用其他部分的音频来补充，从而破坏了其中的隐藏信息。删除的音频可能仅仅只有几十毫秒，这对原始的音频可能是微不足道的一段，但是足以改变隐藏信息的内容。

当删除纯音乐中8000分之一的片段、古典音乐500分之一的片段，人是难以觉察的。使用更加复杂的重采样方法和滤波算法，可以在纯音乐中存在500分之一的抖动、在古典音乐中存在50分之一的抖动的情况下，隐藏信息不存在明显差异。

operation in a transform space在转换空间的操作

上述方法存在一个系统性的问题：隐藏信息可以安全的嵌入在cover filed中很大一个原因是他们都是冗余信息，从而攻击者不会轻易的发现他们的存在，但是这样同样伴随着一个问题，这些冗余信息在有损压缩下会被损坏。因此在学术界，压缩和隐写是密不可分的。

当我们事先直到要采用什么压缩方法的话，可以专门设计一个嵌入方法，从而在该压缩方法下得到较好的效果。比如在GIF文件中嵌入信息，那么可以通过在cover image的大片相似颜色的区域中交换颜色来实现；如果在JPEG文件中嵌入信息，可以通过在多处重复嵌入相同信息来保护信息的完整性、或通过改变图像的离散余弦变换来将隐藏信息嵌入图像的频域。

类似的这种扩频方法是根据cover field的属性决定的。比如可以根据人耳听觉系统的屏蔽属性嵌入信息等。

这种屏蔽属性是一种声音干扰人们对另一种声音感知的现象——当两个频率上接近的音调同时进行播放，较大的声音会影响人们对较小的声音的感知。但是在两个声音的频率存在较大差距时，这种现象不会出现。

另一种现象是当纯音乐被宽频噪声掩盖时，只有纯音乐中心频段的小范围声音会出现屏蔽现象，与之类似的还有当低频信号在高频信号播放前或后立即播放，屏蔽现象就会发生。比如在听到较大的声音后，立即播放音量非常低的声音，我们对这样的声音感知就不会那么强（有时甚至需要休息一会才能听到低音量的声音）。

MPEG音频压缩技术利用了这些特征，但仍有可能通过插入刚好高于MPEG截断阈值但仍低于感知阈值的标记来进一步利用它们。一般来说，版权标志的存在或许可以通过统计测试检测到，但它仍然无法被人类所感知到。对于版权标记而言，真正的问题是它是否能够在不引起可察觉的失真的情况下被破坏到事后无法识别的程度（换句话说就是能否在将版权标志破坏到无法还原的情况后，此时的文件对于人而言与原始文件毫无差别）。

在变换后的内容中嵌入数据并不仅仅是这些“明显”的变换，这些变换被广泛的应用在压缩领域，比如离散余弦变换、小波变换、傅里叶变换等。

a general model通用的模型

上述章节提到的隐写技术的一般模型是：

1. 将cover field进行变换；
2. 调整变换对象比特集的子集，使其变成冗余信息；
3. 在冗余信息内嵌入隐藏信息。

在这种方法中，对于cover field而言，冗余信息是非常重要的，它的大小是不固定的，可以随意更改冗余信息的内容且不会被攻击者轻易检测到，或者攻击者不知道检测哪一个子集。

简单来说冗余信息的量不会太大，因为这样会大大加大通信负担。而压缩方法的设计初衷就是因为经济的原因而用来节省带宽。所以添加的信息只会比正常的信息多一些，从而可以利用较低带宽的通道进行传输。

把高科技作为防止盗版或版权盗窃的手段是错误的，永远不要低估攻击者的技术能力。

为了保证加密尽可能的安全，最好的方法就是使用一次性密钥（香农定理（1949年提出）为了保证加密系统的安全性，密钥的数量要和明文数量一样多（也就是一个明文使用一个新的密钥），这样的话加密手段是绝对安全的，与对手的计算能力无关）(1992年提出了A survey of information authentication一文，关于安全的身份认证提出了一个非常完美的理论，该理论以被应用于核武器控制）。

theoretical limits

我们能得到一个可以提供无条件隐写方法吗？就像一次性便签提供的无条件安全性一样？

假设Alice将一个无压缩的数字视频信号作为cover text，然后将密文以非常低的比率进行编码。比如将密文的第k位嵌入视频中第k帧的某一个像素的最低有效位，像素的选择根据共享的一次性便签的第k个字符控制。

根据直觉，我们认为密文混合在大量的视频原有噪声中（因为编码比例太低了，所以密文嵌入所造成的影响在大量的噪声中显得很不明显），不会被攻击到，但是如何直观的证明这一想法？

首先需要知道证明的标准是什么。一个比较普遍的标准是嵌入信息的隐写文件与原始文件不能攻击者被区分，除非他知道密钥。

与密码学类似，假设攻击者是一个可以计算概率多项式的图灵机，在无条件安全的前提下，假设其可以使用所有可能的密钥进行计算。此时，攻击者可以看到实际嵌入的信息，所以隐写方法必须保证根据任何给定的隐写文件中都可以生成足够可信的嵌入信息，且这些信息不能被任何方法在stego text和cover text之间相互转换。

这一点实现很简单，但是隐写技术比加密技术更难实现的根本原因是，隐写技术完全依赖于cover text的模型。

what if prefect compression existed?如果存在完美的压缩方法会怎样？

信息论人员假设任何的信息都可以被压缩，除非他们不包含任何的冗余信息。这一假设在证明渐进边界和容量时非常有用，但是在隐写领域则显得非常奇怪。

假设存在一个完美的编码方式，其可以对特定的数字媒体类型进行压缩和解压。完全有效压缩意味着被压缩的对象将在相同长度的 bit string集合中密集的存在。因此Alice可以任意获取她想隐藏的密文消息，并通过解压缩方法对其进行操作。结果可能是一个正常的音频记录，视频或任何东西。

上述并不是一个严谨的证明。可以想象，假如一个设备可以将一个长度为n的随机比特字符集压缩成一个概率多项式为1/n的特定类型的对象。这将引起使用众多信息论的结果去验证，同时否定上述的观点。尽管如此，它指明了当处理隐写系统时，那些经典的信息论理论并不会为我们服务。并且，只有在低效压缩时，信息隐写才可以使用，如果存在一个高效的压缩方法，信息隐写的意义就不大了，或者根本不能实现。

entropy熵

假设嵌入信息和随机数据不可区分的（随机信息是完全加密的），此时隐写文件S的信息熵等于cover text的信息熵C与嵌入信息的熵E之和。

因此，为了使我们的嵌入过程不受对手（仅仅检测是否含有隐藏信息）的干扰，我们似乎有两种选择：

1. 保证在对手使用任何方法衡量H©时，H(E)都显得微不足道；
2. 找到一些方法处理cover text从而在嵌入信息之前减小H©的值。例如在嵌入信息之前，使用有损压缩或消除噪声的方法消除C中的一些非必要信息。

因为无法预估对手对处理熵、区分信号和噪声有多么擅长、总会遇到一个能力未知的对手，所以需要尽可能的证明方法的安全性。

但是有一点，当对手获取的隐写文件越多，他就越能发现隐藏的信息具体是什么，这样的话为了保证安全，嵌入信息的比例就会越来越小，甚至趋近于0（也就是说隐写技术在此时已经不可用了）。

然而根据经验事实，密文可以嵌入的通道的确存在，而且人们已经研究了多种类型内容的相关性。因此，在一些应用场景下使用熵进行衡量是有用的。

但是，我们是否真的可以设计一种隐写技术，使其可以抵挡任何人的攻击？

selection channel选择信道

这个想法受香农证明他第二个编码理论的修正信道的启发（这是一个既能发送又可以接收信号的人通知接收者哪一个bit需要调整的信道，它会产生大量噪声和错误校验边界）。

同样，（在囚犯问题中）Alice和Bob使用一个共享的一次性便签决定哪一个cover text中的bit将会存储下一个密文bit，此时的一次性便签可以看作是一个选择信道。如果监狱长的计算能力是无限制的，他可以尝试所有可能的便签信息（其中当然包括正确的哪个便签的信息），所以便签产生的可信密文的数量一定要足够大，这样的话监狱长才不会发现Alice发送的是隐写信息而不是无害的信息（个人认为可以理解为，如果一个人总是说真话，一旦其说了假话就会脸红，就会被发现；而如果一个人总是说假话，他再次说假话时也很难被发现（因为没有什么明显的特征，不会脸红hhh））。

现在回想一下“书密码”或许是有用的——发送方和接收方共享一本书，并将信息加密为一系列指向单词的指针，比如“78216”可能就表示该书第78页第2段的第16个字。只要攻击者不知道正在使用的是哪本书，并且注意不要重用单词(或与之足够接近的单词)，那么”书密码“就是安全的。“书密码”是一种选择信道，虽然计算模型可能看起来有所不同——“书密码”从书开始，然后生成密文；而在隐写系统中，我们从要嵌入信息开始，然后创建隐写文件。在选择通道是真正随机的(使用一次性便签)的情况下，它们是相同的，因为任意的信息都可以嵌入到足够长的任意的cover text中。

一本重复使用的书其可以隐藏信息的容量会越来越低，因为我们可以使用的单词的比例越来越低，主要原因是它存在根据之前的信息进行破解的风险。类似地，如果隐写系统中使用的隐写文件具有不正常的统计数据(例如0和1的不相等数量)，那么它的隐写容量就会更低，因为只有一小部分可用嵌入位置看起来足够随机。

the power of parity 奇偶校验的作用

前面提到过一种方法，为密文bit提供多个位置嵌入，然后再候选位置中去除嵌入后会非常明显的位置（比如在图像中嵌入，要避免那种大面积相同颜色的区域）。

选择信道方法提供了更好的方法取选择合适的嵌入位置——使用一次性便签（或密钥流）选择一组像素，然后嵌入密文bit作为它们的奇偶校验，这样的话信息就可以通过改变最不引人注意的像素来隐藏。

从奇偶校验的角度来看，如果cover text中的每个bit均为1的概率为0.6的话，每个bit对都为1的奇偶校验为0.52。对于三元组而言，奇偶校验为1的概率为0.504。因此通过将每一个嵌入bit编码为隐写文件中奇偶校验为k的bit，可以有效减少嵌入过程对隐写文件统计量也引起的影响，并使其低于所有的阈值。因为这个改变是几何的，所以实际并不需要把k增加特别多。

有一个有趣的权衡：选择信道中的bit越多（即k的值越大），可以在cover text中隐藏的信息越多。在实际情况下，选择信道通常是一个伪随机数生成器，因此可以从其中得到任意数量的bit。

这个方法同样具有很多的限制。例如，假设有一组可用的cover text M(我们可能使用一个新闻机构的封面;我们必须报告事件的一个合理真实的版本，并传送实际发生的事件的照片——也许是稍微篡改过的)。还假设存在一组可用的的编码E，并且每个隐藏位都通过编码规则的选择嵌入位置(如上面描述的方法中的奇偶校验)。那么隐蔽容量将不超过H(E)-H(M)。但这只给了我们一个上界，它没有给我们提供有用的信息，比如告诉我们有多少信息可以安全地隐藏起来。

equivalence classes 等价类

假设Alice使用一个密钥加密散列函数，从而从文档的每一个句子中获取一个bit，她或许可以设置一个macro（宏指令）来检查每一个句子——一旦哈希函数的输出与她希望的下一个嵌入bit不相等，就会哔哔哔的警报。这个警报器可能在每一个句子后面想起，因此Alice可以对错误进行修改。

如果Alice只是使用标准同义词对（如 is able can等），显然这样的话不会引起统计数据的改变从而引起监狱长的注意。这是一个公开的问题——计算机是否可以以一种人类察觉不到的方式替换自然语言文本？也就是是否可以利用同义词替换实现隐写的操作？这还是一个有待解决的问题。反过来，想要设计一个程序检测人类手动实现的隐写文件可能更加困难。

使用同义词对嵌入消息进行编码，是使用信息等价类的一种特殊情况。这些也可以在其他应用中自然出现，例如，当根据大范围地图制作小范围地图时，便需要做出许多关于哪些细节不需要显示的武断决定，特别是处理一些形状不规则的海岸线。此外，当软件被编写时，它包含“胎记”，例如被寄存器存入和弹出的顺序，这些“胎记”曾被IBM用于对软件盗版者的诉讼。

信息的等价类与压缩也是有紧密关系的。如果cover text C1的效果或含义和cover text C2等价，那么压缩算法只需要在这个等价类中选择一个作为代表即可。然而，当C1和C2并不相等时，再使用这个压缩方法的话会造成信息的损失，成为有损压缩。我们再次发现了选择信道的边界信息，可以用来界定有损压缩和无损压缩，但这仍旧只是一个边界，而不是安全边界，对抵抗具有强大能力的对手不会其任何作用。

必须指出，并非所有的隐写技术都会涉及等价类。我们是有可能创造出一系列的图像，相邻的图像之间的差别只有难以察觉的程度，但是开始和最终的图像是明显不同的。对于囚犯问题而言，这和监狱长的检测误差有关，他只容忍一定范围内的细微改变（比如噪声等），一旦超出这个范围，即便没有完全确凿的证据，囚犯也会被抓起来严格审问（比如在战争阶段，允许士兵给家里打电话，所以士兵给国内的家人打电话完全没有问题，某天一个士兵打了海外电话，无论谈话内容是什么，该士兵都会被抓起来审问）。

纯粹主义者可能会得出这样的结论:只有当爱丽丝使用一个真正意义上难以察觉的信道时，Alice才可以确定监狱长不会发现她隐藏的信息。

active and passive wardens主动和被动的监视

上面已经指出了，典型的隐写系统攻击方法包括检测出嵌入内容的存在（注意是仅仅存在，而不是内容），而这种方法对嵌入的版权信息是无效的（谁都知道里面一定有版权标记，对版权标记嵌入的有效攻击应该是对嵌入信息的消除或更改）。

被动的监视仅仅是检测到异常的流量信息，然后向系统外的某个进程发送信息（仅仅是有异常信息存在的信息而已，不关注内容）；主动的监视的目的是尽可能的去除流量中的隐藏信息。在典型的系统中，监视可以分为主动和被动，而在标记系统中，主要关注的是一些软件盗版者（因为他们需要消除或更改标记信息）。

对于软件代码而言，现在软件的“胎记”已经可以在法庭上作为认证作者身份的证据。这些“胎记”在手动编写代码的过程中会或多或少的存在（个人感觉是一些个人的代码编写习惯，如格式，变量、函数的命名等，每个人都有自己的风格），而现在一些编译器自动生成的代码就都千篇一律了，所以在软件制作过程中要慎用它们。解决这个问题的一种方法是故意破坏目标代码，英特尔使用自动、随机的代码片段替换等价的代码片段来生成具有安全性的代码。

可以想象的是，软件创作者和盗版者之间存在一个竞争——谁能在不破坏性能的前提下对代码“破坏”的更厉害。如果软件创作者拥有更好的编译器，那么他在代码中隐藏的信息将不会被盗版者发现；而如果盗版者知道创作者所有的等价代码，那么他可以完全的阻断隐写通道。一般来说，如果一个主动监视者的通信模型和通信双方一样好，那么它可以将所有的隐藏信息提取出来，并可以将所有的隐藏信息使用噪声进行替代。

多数情况下，隐写通道的选择与cover field高度相关。在gif文件和数字化语音中可以隐藏信息的大小已经被测量过了。这里的关键是，在囚犯问题中，如果Alice可以在不影响图像质量的前提下额外嵌入X%的噪声的话，监狱长同样可以，但是当噪声信号和图像原本的信号不容易进行区分时，两者的数据都被限制在根据某些规则或随机选择的嵌入bit上，如果Alice想要防止监狱长发现嵌入信息的存在，可以根据密钥流来选择嵌入bit的位置，而监狱长想要发现嵌入信息的存在则要通过添加他自己的随机标记来缩小信道带宽——这一操作已在操作系统的隐藏信道中实现。

带宽的限制对那些将每个密文bit嵌入多个隐写文件bit作为奇偶校验的系统非常有效，当面对主动监视的盗版者时，软件创作者嵌入信息使用的隐写文件bit越多，盗版者可以在隐写文件中嵌入的噪声信息也越多（也就是创作者可以隐藏版权信息的地方，攻击者也可以嵌入信息）。

public key steganography公钥隐写技术

直到如今（此处指1998年），很多人认为面对一个能力强且有动机的对手，隐写技术要求在通信双方之间事先确定一个共享的密钥，这样的话可以决定在那些bit发生了变化（也就是在那些bit位置嵌入信息）。

with a passive warden 面对被动的监视的对手

在之前的一篇文章中，谈到了面对被动监视，公钥隐写技术是可行的。给定一个可以嵌入任何密文的cover text，这通常又有一个比例——最多嵌入多少密文而不会引起监视者的注意。

假设Alice可以修改cover text中每k个bit之中的至少一个bit（一般是将隐藏信息伪装成这连续k个bit的奇偶校验信息），这意味着监狱长不能从连续k个bit的随机噪声中通过奇偶检验发现异常，因此Alice可以在奇偶校验中编码一个伪随机字符串。

这种伪随机材料随处可见，每个人都可以看到它。所以监狱长不能根据随机测试分辨出隐写文本和真实的cover text之间的区别，而一个合适的奇偶校验函数可以从任何能够嵌入数据的cover text中提取出伪随机数据。

现在假设Alice和Bob在被监禁之前没有机会就密钥达成一致，但是Bob有一个Alice知道的公钥。她可以用他的公钥加密她的秘密信息，并将其嵌入为连续块的奇偶校验。然后，每个可能的接收者将简单地尝试解密他看到的每条消息，但只有Bob会成功（因为使用Bob的公钥进行的加密）。在实践中，用公钥加密的值可以是由会话密钥和一些填充值组成的控制块，而会话密钥将驱动传统的隐写方案。

普通公钥密码学意味着用户可以在之前没有共享密钥的情况下实现秘密通信；我们对公钥隐写术的实现和测试表明，它们也可以实现秘密通信(如果对于已经共享一个秘密的人来说这是完全可能的)。

with a active warden 面对一个主动监视的对手

上面残留的一个问题就是，当面对一个主动监视的监狱长，公钥隐写术是否仍然可以奏效。这种情况下，原本的机构已经不能达到避免发现要求，因为监狱长同样可以对每k个bit的内容进行调整（也就是前面说的——当面对主动监视的盗版者时，软件创作者嵌入信息使用的隐写文件bit越多，盗版者可以在隐写文件中嵌入的噪声信息也越多，你嵌入信息我也嵌入，这样的话你嵌入的信息就会被覆盖、篡改），他甚至可以把每一个bit的奇偶校验位的值置为0。

现在提出一个新的方法，当囚犯问题的模板进行细微的改变，面对主动监视的监狱长，隐蔽的公钥通信仍然可以奏效。假设Alice将发送给Bob的隐写信息也发送给其他接收者，还假设Alice和监狱长均只能修改少量信息（也就是意味着都只能嵌入少量的信息而不影响原本文件的表现，监狱长同样可以对所有可嵌入信息的位置进行修改，从而可能对Alice嵌入的信息造成破坏），Alice可以选择一个短的一次性密钥，从而在cover text中选择合适的bit位置进行嵌入，并且她可以将隐写信息伪装成连续k个bit的奇偶校验信息。如果监狱长怀疑Alice使用这种方法隐藏信息，为了破坏传递的信息，将所有连续k个bit中的一个bit的值进行改写（因为他直到Alice使用的一次性密钥的具体内容，所以这是当前最好的方法）。这样的话可以篡改Alice嵌入的大部分信息，但是大约还有1/e（原文中没有说e的值是多少。。。）个k元组不受影响，根据一个合适的错误校验码，Alice仍然可以通过Bob的公钥加密信息然后进行传输。

一旦Bob收到消息，Alice就广播她的一次性短密钥。Bob现在将这个密钥应用于他所知道中的所有消息仔细进行检查，其中一条消息产生了一个bit字符串，他可以使用自己的私钥对其解密。监狱长现在知道他从Alice那里转发给Bob的一条消息中包含可疑内容，即一个带有错误纠正代码的随机字符串，但是，他不能确定消息是专门指向Bob的，因为他不拥有Bob的私钥。

我们稍微改变了我们的模型，因为我们现在假设Alice可以完整地发送短的伪随机消息。否则，当监狱长看到带有一次性密钥的消息并意识到它的重要性时，他会破坏它或拒绝转发它。然而，在某些情况下，他可能无法做到这一点。例如，Alice可能是自由的，而Bob在监狱里；监狱长也许可以通过监狱的新闻服务器审查爱丽丝发布的新闻，但不能审查她不时在电台节目上发表的评论。