关于 RFID 标签的安全策略研究

宋合营,赵会群
SONG He-ying,ZHAO Hui-qun

北方工业大学 信息工程学院,北京 100041
School of Information and Engineering,North China University of Technology,Beijing 100041,China
E-mail:songheying@gmail.com

SONG He-ying,ZHAO Hui-qun.Study on strategy of security about RFID tags.Computer Engineering and Applications,2008,44(9):109-112.

Abstract: Based on the analysis of previous security strategy about RFID tags,this paper proposes a kind of new security strategy built on random update process which can solve the security problems effectively occurred in the RFID tags,and does not required much computational power and cryptographic techniques.The strategy can change the tag data at random driven by the internal circuitry during each activation and informs the change to the system.Then the system may modify the appropriate entry in its database.After a certain number of activations the new tag data cannot be linked with the original one so as to avoid illegal tracing.Besides,the definition of unlinkability is given,and the detail analysis is provided based on relative theories.
Key words: RFID tags;update process;unlinkability

摘 要: 在比较分析现有的 RFID 标签安全策略基础上,提出一种新型的基于随机更新过程的安全策略。该策略能有效解决 RFID 标签信息的安全问题,并且不需要高强度运算和加密技术。该策略在每次标签激活时由标签内电路随机改变其数据并向应用系统发出通知。应用系统接收到通知后更新已注册的标签数据。经过多次激活之后,标签新数据与旧数据完全不同,从而防止了标签非法跟踪。介绍不可链接性的定义,并在相关理论的基础上给出详细的分析。
关键词: RFID 标签;更新过程;不可链接性
文章编号: 1002-8331(2008)09-0109-04 **文献标识码:** A **中图分类号:** TP391

1 引言

RFID(Radio Frequency Identification,射频识别)安全问题被业界关注。RFID 本身的众多优势使其应用前景广阔(如全球性的物联网有望引领物流领域的变革)。所以,其安全问题不容忽视。

RFID 系统的安全问题涉及前端无线装置和协议、后台网络系统安全以及数据安全等多个方面。本文只关注标签的安全性。目前引发标签安全的隐患有标签克隆、未授权的读写操作、数据的删除、拒绝服务(DoS)攻击、标签病毒等[6]。而最受关注的标签安全问题是涉及个人隐私标签数据的非法跟踪。设想一个包含静态二进制数据的标签,如果没有对标签实施任何的保护机制,则任何读取器都可以访问标签中的数据,此种情况存在于目前大多数 RFID 系统中。对于要求高安全性的用户来说,肯定不会使用这种标签[10]。

RFID 技术安全策略实施的难度有:标签很小,因此在技术上来说,很难给它们提供保护;RFID 标签是移动的,因此可以接触到它的人很多,而且大部分是未授权的用户;标签上的信息并不总是敏感信息,有些情况下,花费太多的时间和费用成本去保证货物 RFID 标签信息的安全性,对于货主来说是毫无意义的;标签的用途非常广,因此在其安全性问题上很难做到标准化和量化。所以,标签数据安全性要求应视其数据敏感程度的不同而不同。对于安全性较高的 RFID 应用(如涉及国家安全、个人隐私、生命财产安全等),需要采用高可靠,抗攻击性的安全策略[11]。

本文提出一种标签数据自更新的安全策略。当多次更新之后,只有合法 RFID 读取器和相应的应用系统可以识别标签数据,非授权者则无法跟踪标签数据。即合法用户与标签是可链接的,而非授权者与标签是不可链接的(不可链接性在第 4 章给出)。

2 现有策略

目前,标签安全的策略有多种,其中最为主要的策略有以下几种:

(1)只读标签:此方式消除了标签数据被篡改和删除的风险,但仍有被非法阅读的风险。

(2)限制标签和阅读器之间的通信距离:采用不同的工作

基金项目: 北京市自然科学基金(the Natural Science Foundation of Beijing City of China under Grant No.4062012);北京市教委科研计划项目资助(No.KM200710009009)。
作者简介: 宋合营(1980-),男,硕士生,主研方向为射频识别、物联网技术、分布式计算;赵会群(1960-),男,博士,教授,研究方向为软件体系结构。
收稿日期: 2007-07-10 **修回日期:** 2007-09-10

频率、天线设计、标签技术和阅读器技术可以限制两者之间的通信距离,降低非法接近和阅读标签的风险,但是这仍然不能解决数据传输的风险,还以损害可部署性为代价。

(3)实现专有的通信协议[5]:在高度安全敏感和互操作性不高的情况下,实现专有通信协议是有效的。它涉及到实现一套非公有的通信协议和加解密方案。基于完善的通信协议和编码方案,可实现较高等级的安全。但是,这样便丧失了与采用工业标准的系统之间的 RFID 数据共享能力。

(4)屏蔽[8]:屏蔽掉标签之后,也同时丧失了 RF 特征。但是在不需要阅读和通信的时候,这也是一个主要的保护手段,特别是包含有金融价值和敏感数据的标签(高端标签,如智能卡)的场合。

(5)使用杀死命令[7](Kill Command):Kill 命令是用来在需要的时候是标签失效的命令。接收到这个命令之后,标签便终止其功能,无法再发射和接收数据。屏蔽和杀死都可以使标签失效,但后者是永久的。特别是在零售场合,基于保护消费者隐私的目的,必须在离开卖场的时候杀死标签。这种方式的最大缺点是影响到反向跟踪,比如退货、维修和服务。因为标签已经无效,相应的信息系统将不能再识别该数据。

(6)物理损坏:物理损坏是指使用物理手段彻底销毁标签,并且不必像杀死命令一样担心是否标签的确失效,但是对一些嵌入的、难以接触的标签则难以做到。

(7)认证和加密[5]:可使用各种认证和加密手段来确保标签和阅读器之间的数据安全。比如,直至阅读器发送一个密码来解锁数据之前,标签的数据一直处于锁定状态。更严格的还可能同时包括认证和加密方案。但是标签的成本直接影响到其计算能力以及采用的算法的强度。因此,在高端 RFID 系统(智能卡)和高价值物品场合,可以采用这种方式。

(8)选择性锁定[7]:这种方法使用一个特殊的称为锁定者(Blocker)的 RFID 标签来模拟无穷的标签的一个子集。这一方法可以阻止非授权的阅读器读取某个标签的子集。这一方法克服或者平衡了以上方法的缺点,也消除了加密和认证方案带来的高成本性。这一方法在安全性和成本之间取得了较好的平衡。需要的时候,Blocker 标签可以防止其他阅读器读取和跟踪其附近的标签,而在需要的时候,则可以取消这种阻止,使标签得以重新生效。

可以说,没有任何一个单一的手段可以彻底保证 RFID 应用的安全。在很多时候,都需要采用综合性的解决方案。本文将为 RFID 应用安全方案的实施提供一个更好的选择。

3 标签数据自更新策略

该策略可实现在非连续访问的情况下,标签对与非授权者是不可链接的。即非授权者获取标签数据后,仍无法获取 RFID 系统内的有效信息。

假设 RFID 标签内存有 n 位二进制数据(为讨论问题方便,n 位数据即为标签 ID,且标签内没有其他任何信息),即 $(b_1,\cdots,b_n)\in\{0,1\}^n$,这里 n 是一个较小的数值,并且每个二进制位都可以被标签内的随机数生成器读写。

下面的步骤精确说明标签数据的更新过程,其中 l 为更新过程的一个参数,表示 n 位二进制数中更新的位数。这里用 $x\in_U S$ 表示从有限集 S 中随机选择元素的值为 x:

(1)从集合$\{1,\cdots,n\}$随机选择 l 个元素组成一个子集 B,即 $B\subseteq\{1,\cdots,n\}$。令 $B=\{i_1,\cdots,i_l\}$。

(2)对于每个 $j\leqslant l$,设置 $b_{ij}:b_{ij}\leftarrow b\in_U\{0,1\}$,即变量 b_{ij} 设置后的值与设置前的值无关。当 l 位的二进制数被随机设置之后,该 l 位的二进制数平均变化 $l/2$ 位(由于每一位的概率满足 0-1 分布,该结论显然成立)。

在每次激活标签时,标签先发送未更新的 ID。RFID 系统接收到数据后,查找数据库,找到与所接收 ID 相同的数据项。接着标签执行数据更新过程并向读取器发送更新后的 ID。RFID 系统接收更新后的数据并进行数据库更新。在异常情况下,标签执行数据更新过程后并没有向读取器发送更新后的数据,则在下次激活标签并发送数据时,RFID 系统需要根据接收到的 ID 在数据库中找到一个匹配项(注,所以接收到的 ID 与系统内部所对应数据项的 ID 的汉明距离 Hamming Distance 不超过 l)。在实际应用中,若新的标签数据发送时遇到干扰或其他原因而导致数据发送失败,就会发生上述异常情况。另外,非授权者极有可能会通过多次读取标签信息以攻击 RFID 系统,此时最简单的办法就是在标签成功通过系统认证之后马上更新标签数据。

这种标签自更新策略不需要高的计算能力,特别是当随机数生成器被外置时(由于更新和发送的信息都是未被加密的明文,所以随机数生成器是否被隐藏起来已无关紧要);该策略的另一个主要的优点是信息无需加密,当数据加密被法律限制时更会显出此策略的优势。此外,具有随机性能的加密技术(如 ElGamal)对系统实现来说可能包含一些潜在信道进行数据隐藏,而这极有可能带来新的安全威胁。对该策略来说,在标签和读取器之间只是简单的明文通讯,这也有利于标签和读取器设备的设计实现。

4 不可链接性

该策略主要关心的问题是:假设非授权者通过一定手段获取一个标签 ID(记为旧 ID),之后标签还要被激活多少次(注:ID 在每次激活后都要更新,且非授权者并没有获取更新之后的 ID),所获得的 ID(记为新 ID)与旧 ID 是不可链接的? 本章将说明安全问题的一个中心概念-不可链接性。

不可链接性的直观定义有多种,每一种之间都有很大差别。适合不可链接性定义如下:

定义 1(不可链接性) 假定非授权者获取了标签 ID(记为旧 ID),之后标签又被合法授权者激活 T 次,数据库种的注册数据项和标签 ID 都发生了 T 次更新。此时利用旧 ID 将不能正确匹配数据库的注册数据项,同时利用旧 ID 也无法找到正确对应的标签,即标签中新 ID 不再与旧 ID 相同,则我们称旧 ID 与新 ID 具有"不可链接性"。

由于策略在每次激活标签时都会更新标签 ID 的若干位,所以若干次标签激活之后,将得到一个随机的 ID。再进一步考虑标签激活 T 次之后所达到的所有 ID 及其概率(因为每次激活标签时新 ID 产生的可能性都有多种)。可能有人倾向于只关心所达到的可能的 ID 数量,也可能对达到 RFID 系统中所有 ID 的 T 比较感兴趣(这里可以考虑类似的息票收藏家问题)。然而,这样考虑对实际应用价值不大,因为可达所有标签数据的概率不一定相同,并且分析结果仅仅只是在一定的概率范围内为用户提供有价值的信息。

考虑在标签激活 T 步之内可能达到的 ID 的概率分布,并且认为此分布在概率空间 $\Omega=\{0,1\}^n$ 内近似服从均匀分布。为比较两个随机变量及其概率分布,本文使用一种总变化距离

(Total Variation Distance,简称 TVD)的度量标准。设随机变量 $\Gamma_1,\Gamma_2:\Omega\to\gamma$,则 Γ_1,Γ_2 的 TVD 表示为:

$$TVD(\Gamma_1,\Gamma_2)=(1/2)\sum_{y\in\gamma}|\Pr(\Gamma_1=y)-\Pr(\Gamma_2=y)|$$

这里,对任意 $y\in\{0,1\}^n$,$\Pr(\Gamma_U=y)=1/2^n$ 办,所以 $\Pr(\Gamma_2=y)$可简化为 $1/2^n$。

定义 2(匿名化) 设 R_t 为一随机变量,表示更新 t 次之后的 ID,R_0 为更新开始之前的旧 ID。第 t_0 次标签数据更新后,称 RFID 标签被匿名化,如果满足对于 $t\geqq t_0$,有

$$TVD(R_U,R_t)<c$$

其中,参数 c 通常为 $1/n^{\alpha}$,$\alpha\geqq 1$。

定义 2 是定义 1 更严格的数学表述,是定义 1 的补充,其优点是基于 TVD 之上定义较之其他文献更为准确可靠,很适合本文讨论。了解更多不可链接性定义可参考文献[3]。文献[11]也给出了其他一些不可链接性的定义。在下面的讨论中,为简化表述,有的随机推导过程的略去一些参数,如用 X_t 代替 $\{X_t(\omega)\}_{t\in N}$。

4.1 主要结论

先给出混合时间的概念:收敛于 X 的遍历有穷马尔可夫链 X_t 的混合时间为:

$$\tau(\varepsilon)=\max_s\min_t\{t\in N|TVD(X_t,X)\leqq\varepsilon\Lambda X_0=s\}$$

这里,混合时间可理解为:马尔可夫链的第一步转换。其前提是,对于初始状态 X_0,在最坏情况下,马尔可夫链变为对于平稳分布的 ε-闭包。换句话说,经过时间长度为混合时间 $\tau(\varepsilon)$之后,当前状态与起始状态将不可链接。

主要结论可用下面的定理来描述:

定理 1 设 R_t 为 RFID 标签在第 t 次激活之后的状态(如何更新已在第 3 章论述),R_t 起始于任意的一个初始状态。在标签不断更新的过程中,对任意 $k>1$,有

$$\tau(1/n^k)\leqq(n*\log n^{k+1})/l$$

所以,当 $n=64$,$l=10$ 时,得到 $\tau(2^{-24})\leqq 64.2$。

4.2 耦合引理

证明定理 1 之前,先给出 3 个引理,提供同类遍历的马尔可夫链收敛率的上限。设 γ_t 为有限状态空间 S 上的离散遍历的马尔可夫链,且服从唯一的平稳分布 μ。

引理 1 马尔可夫链 Y_t 的耦合是一个随机过程 $(Y_t,Y_t^*)\in S\times S$。其中 Y_t 和,Y_t^* 分别为 γ_t 的完全拷贝。例如:对任意 $x,y\in S$,有:

$$\Pr(\gamma_{t+1}=y|\gamma_t=x)=\Pr(Y_{t+1}=y|Y_t=x)=\Pr(Y_{t+1}^*=y|Y_t^*=x)$$

这里,过程 Y_t 和 Y_t^* 可随机依赖。

下面给出证明定理 1 需要使用的一个关键方法:

引理 2 设 $Z_t=(Y_t,Y_t^*)$为空间 S 上遍历马尔可夫链的一个耦合。若对任意 $x,y\in S$,有,$\Pr(Y_t\neq Y_t^*|Y_0=x,Y_0^*=y)\leqq\varepsilon$,则

$$\tau(\varepsilon)\leqq T$$

此引理的证明参考文献[1],[4]。

在定理证明中也使用到了下面的息票收藏家引理:

引理 3 考虑向盒子里放球的实验,设有 n 个盒子,每次都从 n 个盒子中随机选择一个盒子(即每次放球之间相互独立),向选定盒子里放一个球。设 C_t 为第 t 次向盒子放球之后空盒的数量,有 $\Pr(C_t\geqq 1)\leqq(1-\frac{1}{n})^t n$。通过求线性期望得到,对任意 $t\geqq 0$,有 $E(C_{t+1}|C_t)=(1-\frac{1}{n})^t E(C_t)$,所以 $E(C_t)=(1-\frac{1}{n})^t E(C_0)=(1-\frac{1}{n})^t n$。由于 C_t 仅有非负值,所以 $\Pr(C_t>0)\leqq E(C_t)\leqq(1-\frac{1}{n})^t n$。

4.3 定理 1 证明

首先,为了表示单个 RFID 标签 ID 的变化,需要定义一个随机过程 R_t。显然,此过程是一个马尔可夫链,因为第 $t+1$ 次变化之后的状态仅依赖第 t 次变化之后的状态。设 $R_0=s$ 为变化过程开始前 RFID 标签的 ID 号。则可得到下面的转移概率表达式:

$$\Pr(Rt+1=y|Rt=x)=\begin{cases}\binom{l}{i}\left(\frac{1}{2^l}\right)\binom{n}{i}^{-1} & \text{对于 } i\leqq l,\text{有 } d_H(x,y)=i \text{ 时}\\ 0 & \text{其他}\end{cases}$$

这里,$d_H(x,y)$是 n-bit 字符串 x 和 y 的汉明距离。显然,R_t 具有非周期性且不可约。因此,R_t 收敛于唯一的平稳分布。容易看出 R_t 是对称的,所以平稳分布是均匀的。

下面开始求均匀分布的收敛率和混合时间 τ 的界。为此需要使用耦合引理。首先是构建耦合。这里为过程 R_t 构建一个耦合(X_t,Y_t)。设 X_t 为起始于任意状态 $x_0\in\{0,1\}^n$。在变换过程的每一步,X_t 根据 R_t 的转移概率进行变换(注:变换的每一步之间相互独立)。

第二个过程 Y_t 起始于一个随机状态 y_0。Y_t 在第 t 步的变换依赖过程 X_t 的变换。即,Y_t 选择一个更新位置,并用 X_t 相应位置的值重写所选位置。

现在可得到两个重要的结论:一是 Y_t 和 R_t 有相同的转移概率,所以(X_t,Y_t)是一个有效耦合;二是当每个比特位都发生变化之前,过程 X_t,Y_t 将达到相同的状态。即,标签可经过多次更新,以使 ID 所有比特位都被选择。

这里,将估计事件$\{X_T=Y_T\}$发生的概率。设 N_t 表示为标签 ID 的前 t 次更新后没有被选择的比特位。显然,$N_0=n$。

考虑若对任意 $x\in R$,随机变量 V_1 随机支配随机变量 V_2,则 $\Pr(V_1>z)\geqq Pr(V_2>x)$。也可以得到随机变量 $C_{k\cdot t}$ 支配随机变量 N_t,即,考虑将 $l\cdot t$ 个球随机放入 n 个盒子后空盒的数量。所以,可根据引理 3 估计事件$(X_T=Y_T)$的概率为:

$$\Pr(X_T\neq Y_T|X_0=x,Y_0=y)<(1-\frac{1}{n})^{l\cdot t}\cdot n$$

根据引理 2,得到混合时间的上界:

$$TVD(R_t,R_U)<(1-\frac{1}{n})^{t\cdot l}\cdot n=((1-\frac{1}{n})^n)^{(t\cdot l)/n}<e^{-t\cdot l}\cdot n$$

上述推导依据数列 $a_n=(1-\frac{1}{n})^n$ 严格递增并收敛于 $1/e$。至此,定理 1 证毕。

5 策略之不足

上述内容已经说明策略可以有效地解决标签安全问题。标签 ID 经过多次更新之后,新 ID 和旧 ID 间的依赖将消失。另外,因为标签 ID 是动态独立地在标签内部更新而不依赖 RFID 应用系统,所以会遇到下面的问题,即在更新过程中两个不同标签 ID 经过更新得到相同地 ID,此时两个标签将发生冲突,因而 RFID 系统将无法识别标签。

解决以上问题有个简单方法。若一个标签更新之后其 ID

很接近其他标签的ID，则系统将多次更新此标签直到标签ID之间的汉明距离达到要求的最小值。下面有两个问题需要解决：(1)经过M步更新之后，RFID系统中标签ID发生冲突的概率是多少？(2)多长时间以后两个标签ID将达到相同的值？

5.1 冲突概率及冲突时间

5.1.1 冲突概率

设 R_t^1，R_t^2 为两个不同标签的更新过程(相关内容前几章已经论述)。定义另一过程 $D_t^{1,2}=R_t^1\oplus R_t^2$。即，给定随机变量 t，$D_t^{1,2}$：$\Omega\to\{0,1\}^n$ 返回 n 比特的字符串；其中，若 R_t^1 和 R_t^2 在位置 i 值不同，则字符串第 i 位为1，否则为0。所以，对过程 $D_t^{1,2}$，有

$$\tau(1/n^{k-1})\leqslant(n\cdot\log n^k)/l \tag{1}$$

该结论的正确性是因为过程 $D_{1,2}=R^1\oplus R^2$ 和上面几节中论述的 R_t 几乎完全一样。仅有的不同是在 l' 位置放入了随机内容，其中 $l\leqslant l'\leqslant 2l$，而不是在 l 位置放入随机内容。事实上，要想改变某一位置，只需过程 R_t^1 和 R_t^2 其中之一选择此位置就可以了。这样，就可以为过程 R_t 重复使用此规则。根据结论(1)和定义 $TVD(\cdot,\cdot)$，可以得到满足 $t\geqslant\tau(\varepsilon)$ 时，两个标签ID达到相同的概率如下式：

$$\Pr(D_t^{1,2}=0^n)\leqslant 1/2^n+\varepsilon$$

下面求集合 s 中的任意两个标签ID在 $\tau(\varepsilon)$ 时间之后达到相同值时的概率（这里RFID应用系统中所有的标签ID为集合 s）。这个概率值用每一对标签冲突的概率乘以可能产生冲突的标签对数来估计。即：

$$s(s+1)/2\cdot(1/2^n+\varepsilon)$$

这里，利用以前的假设 $\varepsilon=1/n^k$。则得到标签ID冲突的最小概率近似为 $s(s-1)/2^{n+1}$。

5.1.2 冲突时间

如果两个标签ID相似，可得到它们发生冲突所需时间。当两个标签达到上述的均匀分布时，冲突很有可能发生。考虑下面的随机过程：有两个标签，每个标签分别有 n 位ID。标签每一步更新，都随机的选择一位 $m\leqslant n$，则标签ID第 m 位可能发生变化。

同样，可考虑对单个标签，看需要经过多长时间此标签ID所有位都为0。对于每个ID，其汉明权重(Hamming Weight)决定其更新之后所有位都为0的概率。即，若标签ID中有 w 个1，那么一次更新之后，到达目标的距离值小1的概率为 $d=w/n$，到达目标的距离大1的概率为 $1-d$。所有当 w 很小时，d 接近于 $1/n$。为了分析到达目标所需要的时间，考虑具有状态1，…，k 的过程 P，其中1为初始状态，k 为目标状态。则状态 $1\leqslant i<k$ 的转移概率如下：当 $i>1$ 时，状态 $i+1$ 的概率为 d，状态 $i-1$ 的概率为 $1-d$；当 $i=1$ 时，状态1的概率为 $1-d$。文献[9]给出了达到状态 k 的期望时间可用下式表示：

$$T=\frac{2}{d}+\frac{k-2}{2d-1}+\frac{(1-d)^2(2d-2)}{d(2d-1)^2}\left(1-\left(\frac{1-d}{d}\right)^{k-2}\right)$$

所以，当 d 远小于1时，T 为 $1/d^k$ 的阶。当然，过程 P 达到状态 k 的期望时间并不大于其达到ID所有位都为0的状态所需的期望时间，因为过程 P 中到达目标的距离会增加并超过初始值。我们能算出在过程 P 中到达目标状态所需要的平均时间：当 $w=\sqrt{n}$ 时，有 $d=1/\sqrt{n}$，所以 $T\approx n^{\sqrt{n}/2}$；当 $w=n^{1/8}$ 时，有 $d=1/n^{7/8}$，所以 $T\approx n^{7/8\cdot n^{1/8}}$。在两种情况下，$T$ 都是关于 n 的复杂多项式，所以到达均匀分布(1)所需要的时间稍微比线性高一些。

结果说明只有当标签ID位数变成和地址空间(即 2^n)相同的阶时标签冲突才会频繁发生。

5.2 非冲突阶段

若两个标签ID达到相同值所需要的时间很长，这将不会引起任何实际的问题，标签也不会发生冲突。但在实际应用中，冲突发生的频率要高的多。所以这里仍需关注RFID系统中的标签冲突时间。

这里，考虑随机映射 $K_{n,k}:\Omega\to(\{0,1\}^n)^k$ 随机生成一个自 $\{0,1\}^n$ 的 k 元组序列，即，对每个 $(x_1,\cdots,x_k)\in(\{0,1\}^n)^k$，有 $\Pr(K_{n,k}=(x_1,\cdots,x_k)=2^{-nk})$。

再考虑下面一个元组 $(x_1,\cdots,x_k)=K_{n,k}(w)$ 到另一个元组 $\kappa_{n,k}^*$ 转换：独立地等概率地随机选择一个元素 $(i,j)\in\{1,\cdots,k\}\times\{1,\cdots,n\}$，然后挑出比特 $x_i(j)$。显然，随机变量 $K_{n,k}$ 和 $\kappa_{n,k}^*$ 具有相同的分布。

考虑序列 $(\kappa_{n,k}^1,\cdots,\kappa_{n,k}^M)$ 到 $(\{0,1\}^n)^k$ 的随机映射，这里 $\kappa_{n,k}^1=K_{n,k}$，且 $\kappa_{n,k}^{i+1}$ 是由 $\kappa_{n,k}^i$ 根据上述转换规则得到的。显然，结果映射 $\kappa_{n,k}^{(i)}$ 是强相关的。设 $L_M=\sum_{i=1}^M 1_{\kappa_{n,k}^i\ \text{is}\ 1-1}$，其中，$1_E$ 表示事件 E 的特征函数。则 $E|L_M|=\sum_{i=1}^M\Pr[\kappa_{n,k}^i\ \text{is}\ 1-1]$ 为更新过程非冲突阶段的期望值(更新过程发生在长为 M 的随机性系列集中)。通过文献[5]可得，若 $k<\sqrt{\pi 2^{n-1}}$，则 $\Pr[\kappa_{n,k}^i\ \text{is}\ 1-1]\simeq 1-e^{-\frac{k^2}{2^{n+1}}}$，所以 $E|L_M|\simeq M(1-e^{-\frac{k^2}{2^n}})\simeq\frac{Mk^2}{2^{n+1}}$。

注意到随机变量 L_M 的值为正整数，所以有 $\Pr[L_M\neq 0]\leqslant E[L_M]$，则序列 $(\kappa_{n,k}^1,\cdots,\kappa_{n,k}^M)$ 至少发生一次冲突的概率的上界为 $\frac{Mk^2}{2^{n+1}}$。

因此，如果 $k<\sqrt{\pi 2^{n-1}}$，$p\in[0,1]$ 并且 $M<\frac{p2^{n+1}}{k^2}$，那么序列 M 至少发生一次冲突的概率小于 p，其中 M 为自 $\{0,1\}^n$ 的 k 序列的随机集合的单个比特连续修改得到的序列。

6 平均最小距离

本章将求解标签ID之间的平均最小距离。这个问题之所以重要，是因为保持最小距离会使RFID系统针对某些传输和硬件故障有较高的容错性。另外当一个标签更新之后，更新的值并没有传给读取器，这时保持标签ID之间的最小距离也很重要。

集合 $C\subseteq\{0,1\}^n$ 的距离定义为：$d(C)=\min_{x,y\in C;x\neq y}d_H(x,y)$，这里 $d_H(x,y)$ 为 x 和 y 之的汉明距离。C 的相对距离(Relative Distance)定义为：$\delta(C)=d(C)/n$。C 的秩(Rank)为：$R(C)=(\text{lb}|C|)/n$。这里给出一个来自编码理论(文献[2])的结论：

定理2 对比率为 R 的 $\{0,1\}^n$ 的所有子集 X(不包含以指数 n 递减的子集)，相对距离的上界为

$$2R=1-h_2(\delta(X)) \tag{2}$$

其中，$h_2(\delta)=-(\delta\text{lb}(\delta)+(1-\delta)\text{lb}(1-\delta))$。

(下转142页)