基于多决策树的 RFID 入侵检测模型*

杨晓明[1,2]，丁　哲[1]，秦志光[1]，王佳昊[1]
（1. 电子科技大学 计算机科学与工程学院，成都 611731；2. 工业和信息化部电子第五研究所，广州 510610）

摘　要：为了能够有效对 RFID 系统的已知攻击和未知攻击进行检测和预警，在分析大量 RFID 安全协议的基础上，提出了利用构建多决策树的方法对 RFID 系统的入侵行为进行检测。通过分析，得到所构造出的多决策树能够发现针对 RFID 系统的攻击之间的关系，并且具有可扩展性和高效性，使之能够适用于各种实际的 RFID 系统的应用。

关键词：RFID；安全隐患；入侵检测；安全协议；多决策树
中图分类号：TP311；TP393.08　　**文献标志码**：A　　**文章编号**：1001-3695(2014)04-1172-04
doi：10.3969/j.issn.1001-3695.2014.04.052

Intrusion detection model based on multi-decision trees for RFID technology

YANG Xiao-ming[1,2], DING Zhe[1], QIN Zhi-guang[1], WANG Jia-hao[1]
(1. *School of Computer Science & Engineering*, *University of Electronic Science & Technology of China*, *Chengdu* 611731, *China*; 2. *The Fifth Research Institute of MII*, *Guangzhou* 510610, *China*)

Abstract: In order to precaution known and unknown attack of RFID technology, this paper proposed multi-decision tree model based on the analysis of RFID security protocol to check the intrusion behavior. By the analysis of the multi-decision tree model, the model can find the relationship of different attack of RFID technology. And the model has the advantage of extensibility and high efficiency, which make the multi-decision tree model suitable for all practical application of RFID system.
Key words: RFID; security risk; intrusion detection; security protocol; multi-decision tree

0　引言

RFID(radio frequency identification)技术即射频识别技术，它是一种利用射频通信实现的非接触式的自动识别技术。由于标签可支持快速读写、非接触识别、高速移动识别、多目标识别、定位以及长期跟踪管理，因此，RFID 技术已经广泛地应用于制造、物流和公共制造等行业[1,2]。

目前基于 RFID 技术的应用设计是完全开放的，因此 RFID 技术在实际应用中缺乏完善的安全机制。而且由于目前市面上绝大多数的 RFID 标签的计算能力比较弱，所以这些 RFID 标签上不支持加、解密功能。标签与阅读器交互的过程是以明文的形式传输的，这就导致 RFID 技术在应用中存在非常严重的安全隐患。虽然后来改进方案将用户的敏感信息在阅读器端进行了加密，并且将密文写入标签，但这并没有从根本上解决这一隐患[2,3]。

当前国际上主要有三大 RFID 标准体系，即 ISO 制定的 RFID 标准体系、EPC Global 制定的 RFID 标准体系、日本 UID 制定的 RFID 标准体系[4]。但这些标准体系存在非常严重的安全漏洞。目前针对 RFID 系统的攻击可以大体分为四类，即 DoS 类、跟踪攻击类、隐私泄露类和伪造数据[5]。近年来很多学者在这三大标准体系的基础上，针对这四类攻击提出了很多新的改进协议，但这些改进协议还是存在很多安全隐患[5~7]。

假设敌手具有对射频信息进行分析、从中得到用户敏感信息的能力，本文在分析三大标准协议和改进协议的基础上提出了利用决策树算法来对 RFID 系统中的攻击进行入侵检测。传统决策树的最大缺点之一是属性的构造。一旦决策树构造出来，如果产生新的问题，需要将新的属性和新的类别考虑到决策树中，那么就必须重新生成训练集，重新构造。本文采取多决策树的方法，这样不但方便后续决策树的扩展，而且可以通过多决策树模型发现类与类之间的关系。本文主要是针对 RFID 标签无加、解密功能时，RFID 系统在认证过程中的针对阅读器与标签交互的非暴力入侵行为进行检测。

1　基于决策树的入侵检测

1.1　决策树的构造

1980 年，Quinlan 在早期 Hunt、Marin 和 Stone 的概念学习系统工作的基础上提出了决策树算法 ID3。后来 Quinlan 又提出了 ID3 的后继算法 C4.5。而 C4.5 成为了新的监督学习算法性能对比的基准。1984 年，统计学者 Breiman、Friedman、Olshen 和 Stone 在《分类和递归树》一书中介绍了二叉决策树的生成[8,9]。

决策树构造采用了贪婪算法，其构造过程采用了自顶向下

收稿日期：2013-07-13；**修回日期**：2013-08-27　　**基金项目**：国家科技重大专项课题(2011ZX03002-002-03)；中央高校基本科研业务费资助项目(ZYGX2011J066)；四川省科技支撑计划资助项目(2013GZ0022)

作者简介：杨晓明(1974-)，男，高级工程师，博士研究生，主要研究方向为信息安全(yangxm@ceprei.com)；丁哲(1982-)，男，博士研究生，主要研究方向为信息安全；秦志光(1956-)，男，教授，博士，主要研究方向为信息和网络安全、电子商务、智能交通系统；王佳昊(1978-)，男，副教授，博士，主要研究方向为 RFID、无线传感器网络.

和分治的方法。其构造过程如下:令 $A=\{a_1,a_2,\cdots,a_n\}$ 为训练集的属性集,$B=\{b_1,b_2,\cdots,b_m\}$ 为决策树的类别集,a_i 的属性值为 $\{c_1,c_2,\cdots,c_k\}$。首先将训练集与它们所属的类别进行关联;其次利用信息增益 gain 在训练集的属性集中选出最优分裂属性。可以通过式(1)~(3)得到信息增益,其中 p_i 是指训练集中属于 b_i 类的元素所占比重;B_j 表示训练集中,在 A 的 a_x 属性中具有 c_v 值的元素的集合[8,9]。图 1 是购买电脑的决策树。

$$\mathrm{info}(B)=-\sum_{i=1}^{m}p_i\times\log_2 p_i \tag{1}$$

$$\mathrm{info}_A(B)=\sum_{j=1}^{v}\frac{|B_j|}{|B|}\times\mathrm{info}(B_j) \tag{2}$$

$$\mathrm{gain}(A)=\mathrm{info}(B)-\mathrm{info}_A(B) \tag{3}$$

根据最优分裂属性的值将训练集划分为若干个子集,然后在每个子集中递归地选取新的最优分裂属性,并将该子集进行分裂,直到无属性划分或最终子集都属于一个类别,如图 1 为购买电脑的决策树。目前决策树分类已经广泛地应用于很多商业软件中,并且事实证明了决策树性能的可靠性。

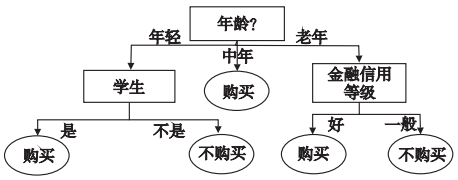

图1　购买电脑决策树

1.2　从决策树上产生分类规则

决策树建立后,通过对决策树进行自顶向下的深度遍历,就可以提取分类规则。在遍历时,从树的根节点到树的叶子节点,每一条路径就表示一个分类规则。在路径上的每一个中间节点都是判断条件,判断条件之间用逻辑与(and)操作连接,最后的叶子节点就是满足前面规则的类别。从图 1 的决策树中可以得到形如:年轻 and 学生→购买电脑的关联规则[10]。

传统的决策树很难判断出类与类之间的规则,在针对 RFID 系统的攻击中,某些攻击可能是其他攻击的组成部分,而传统的决策树很难发现这种攻击与攻击之间的关系。本文通过建立多决策树,利用其得到的关联规则发现攻击之间的关系。

2　RFID 入侵行为采集

2.1　RFID 标准协议的认证过程

目前国际上有三大标准协议,这些标准协议的 RFID 系统的认证过程一般分为以下四步:

a)阅读器向标签发送请求。

b)标签收到请求后,会将自己的身份信息和随机数发给阅读器。

c)阅读器根据标签发来的身份信息对标签进行认证。如果对标签身份认证通过,则将标签发来的随机数再发给标签。

d)标签通过对接收到的随机数进行比对,完成标签对阅读器的认证。

由于在此过程中标签与阅读器之间的射频交互是以明文形式进行的,因此,攻击者非常容易对其进行攻击。所以对于安全性要求高的 RFID 系统,这些标准协议是无法满足其要求的[2,4]。

2.2　决策树的属性定义

由于 RFID 标准协议存在很大的安全隐患,所以研究者在这些标准协议的基础上提出了很多非标准协议。为了建立对于 RFID 入侵检测的决策树,首先从以下几个方面建立决策树的属性:

a)阅读器向标签发送请求,标签可以判断发送该请求的阅读器是否合法(legal reader)。在 RFID 认证的第一步,如果标签无法判断出发送请求的阅读器是否合法,那么攻击者就会非常容易地欺骗标签与它进行交互[5,6]。

b)阅读器收到标签的反馈信息时,阅读器是否可以判断这个标签为合法标签(legal tag)。如果阅读器无法判断,则系统容易受到重放攻击[5,11]。

c)标签判断阅读器向标签发送的请求是否特殊(checking request)。为了保证 RFID 认证过程第二步的安全性,目前大部分的非标准协议中,标签收到阅读器发送过来的请求后,都会对其进行处理,例如将请求和自身生成的随机数进行椭圆运算。攻击者可以利用这些运算的特殊性,假冒阅读器发送特殊的请求,就会导致 RFID 系统的敏感信息泄露[5]。

d)标签在收到来自阅读器的请求后进行何种运算处理,并且,标签生成的随机数是否参与运算(tag processing)。目前大部分的非标准协议为了保证 RFID 认证第二步的安全性,在标签收到来自阅读器的请求后,会对其进行运算处理,如椭圆运算和 hash 处理,并且在这些处理过程中,自身生成的随机数也会参与[5,12,13]。

e)在 RFID 认证的第三步,阅读器收到标签返回的信息后,检测自己向标签发送的请求标签是否收到(checking receiving request of tag)。因为在阅读器与标签认证时,攻击者可以利用中间人攻击的方法对请求进行窜改。为了防止这一攻击,很多 RFID 非标准协议采取阅读器对标签传来的数据进行检测,以确定自身的请求标签是否收到[5]。

f)一次 RFID 认证过程中,阅读器和标签的同步信息是否升级,并且以何种方式升级(updating method of synchronization information)。如果 RFID 标签的敏感信息不升级,那么在每次认证过程中标签与阅读器交互的敏感信息都是固定的,这就导致了攻击者会利用这一点对标签的携带者进行跟踪。为了减少标签携带者被跟踪的可能性,目前的一个方案就是在每次交互以后对标签和阅读器的关键数据进行升级。这种升级方式可以是标签和阅读器相互独立地升级或者是标签和阅读器依赖于所受到的数据进行升级[5]。

g)一次 RFID 认证过程中,阅读器和标签的同步信息如果升级,并且标签的升级,运算中包括阅读器传来的数据,那么标签是否检查所收到的阅读器所发来的升级信息,是否被非法修改(checking updating data)。因为攻击者可以修改升级信息,造成系统的去同步[5,14]。

h)一次 RFID 认证后,阅读器与标签的同步数据是否一致(checking synchronization information)。在 RFID 的认证过程中,阅读器和标签都会进行相互认证,而认证的关键之一就是同步数据的对比。攻击者可以利用这一点,使得标签与阅读器的同步数据发生不一致,从而导致阅读器对标签的拒绝服务攻击[5,15]。

i)监听阅读器与标签的一次认证过程,从交互的数据中抽取出关键数据进行有限操作,然后将操作的数据发给标签,标

签返回的数据是否含有一定的规律(replaying result of sensitive information),例如相等。由于在 RFID 认证过程中,标签自身只能进行一些相对简单的运算处理,这就导致将一次认证的关键数据进行简单处理,再将其发送给不在认证状态的标签,如果标签返回的数据是可以提前预知的,从而用户的敏感信息被泄露。这就会被攻击者利用,从而对标签携带者在标签敏感数据下次升级前进行跟踪[5,16,17]。

j)在 RFID 的一次交互过程中,拦截阅读器发往标签的写入数据,将其修改,并写入标签。然后向标签发送请求,检查标签回复的信息是否含有一定的规律(check modified data)[5,16],例如相等。

k)RFID 的一次认证过程中,用户敏感信息的传输形式(the sensitive information form of the user in radio frequency transmission),如明文或密文。

l)在 RFID 标签与阅读器进行认证的过程中,如果其他阅读器向此标签发送请求,标签的同步数据是否发生改变(checking synchronization information of tag receiving the others request)。如果标签的升级是独立的,那么当标签与阅读器进行交互时,攻击者就会利用其他阅读器向标签发送请求,改变标签的同步数据,从而使得这个交互过程中发生去同步事件[5]。

m)RFID 的安全协议在认证过程中是否会泄露用户敏感信息的部分位(leaking bit information of sensitive information)。在 RFID 交互过程中,如果每次交互时都会泄露部分位的信息,那么攻击者就会对没有泄露的位信息进行暴力攻击[5,17]。

n)在每次 RFID 认证过程中,信道中传输的标签向阅读器发送的认证信息是否服从概率分布,且服从何种分布(authentication information obeying probability distribution)。在此过程中攻击者就有可能从以前的信息和其服从的分布,推测出以后标签向阅读器发送的认证信息[5]。

o)是否能够从 RFID 认证交互中阅读器向标签写入的数据中判断出下一次认证交互中标签向阅读器发送的认证信息(getting authentication information of tag)。

以上是关于构造决策树的属性选取的准则。具体的选择要根据具体的 RFID 系统所选取的安全协议进行选取,从而得到精确度高且结构相对简单的决策树。

2.3 提取决策树的训练集

在基于决策树的入侵检测模型中,首先根据具体的 RFID 系统的日志和属性的选取,产生构造决策树的训练集。而这些训练集不仅反映 RFID 系统的运行情况,而且还反映 RFID 系统在受到攻击时的运行情况。为了能够制造分类器,必须将这些训练集与所属的攻击类别进行关联。然后将这些训练集分为四个子集。每个子集包括正常运行的数据和产生同一类攻击的数据。然后分别为每个子集构造决策树。

本文选取三个属性为例,构造关于 RFID 跟踪攻击的决策树。通过对 RFID 系统的攻击行为日志和正常运行日志的分析,得到决策树的训练集,训练集如表 1 所示,其中变量 *CR*、*TP*、*CRT*、*RSI* 分别代表 Tag 收到的请求是否是特殊值、Tag 的运算方式、一次交互后同步数据是否一致和监听阅读器与标签的一次认证过程,从交互的数据中抽取出关键数据进行有限操作,然后将操作的数据发给标签,标签返回的数据是否含有一定的规律。Result 表示运行结果是攻击成功。*TP* = 0 表示 Tag 中的操作使用近似代数中的群运算;*TP* = 1 表示 Tag 中的操作使用近似代数中的群运算;且 Tag 自身不产生随机数或随机数不作为群运算输入值;*TP* = 2 表示 Tag 中的操作使用 hash 函数,且 Tag 自身不产生随机数或随机数不作为 hash 函数的输入值。*TP* = 3 表示 Tag 中的操作使用 hash 函数,且 Tag 自身随机数作为 hash 函数的输入值。

表 1 RFID 跟踪攻击决策树训练集

ID	*CR*	*TP*	*CSI*	result	ID	*CR*	*TP*	*CSI*	result
1	是	0	一致	是	5	否	2	不一致	是
2	否	0	一致	否	6	否	2	一致	否
3	否	1	一致	否	7	是	3	一致	否
4	是	1	一致	是	8	否	3	不一致	否

通过训练集得到决策树,决策树如图 2 所示。对于所得到的决策树,可以利用自顶向下的方式遍历决策树,从而得到关联规则。如 IF *CR* = 是 and *TP* = 1 THEN 攻击成功。

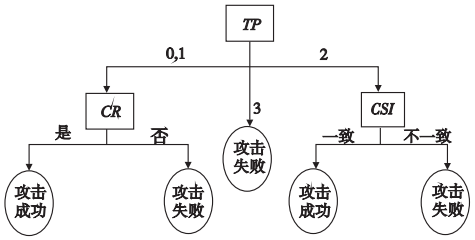

图2 跟踪攻击决策树

3 多决策树的构造

目前针对 RFID 系统的攻击可以分为四类,所以可以同时构造出四个决策树。每个决策树分别对一类攻击进行入侵检测。由于 RFID 系统的特殊性,使得针对 RFID 系统的攻击存在相互关联,如跟踪攻击可以作为 DoS 攻击的后续攻击,并且从攻击的结果分析,有的攻击同时属于两类攻击。为了准确得到系统的安全隐患,本文通过分析多决策树之间的链接关系来对 RFID 攻击进行分析。

在构造完所有的决策树后,将每个决策树的叶子节点的父节点与其他决策树的相同属性的节点相连接。假设构造出的决策树的集合为 $F = \{f_1, f_2, f_3, f_4\}$,首先获得在一个决策树 f_1 中攻击成功的路径 r_1,然后通过叶子节点的父节点,设为 v_1 的链接,可以找到父节点 v_1 在决策树 f_2 的位置。遍历决策树 f_2 的根节点到攻击成功的叶节点,且包含父节点 v_1 的路径(需要考虑到分裂节点的数值输出)r_2。如果 r_1 包含 r_2(不考虑路径中节点的位置),则决策树 f_1 所代表的攻击在满足 r_1 的条件下,是决策树 f_2 所代表的攻击的一部分。如果 $r_1 = r_2$,则 r_2 能引起决策树 f_1 所代表的攻击和决策树 f_2 所代表的攻击这两类攻击。如图 3 所示,左边的决策树是 RFID 跟踪攻击的决策树,右边的决策树是判断 RFID 的 DoS 攻击的决策树。通过跟踪攻击的决策树,可以得到路径:DoS 攻击在满足 *TP* = 2 的条件下是跟踪攻击的一个步骤。

上述是针对两个攻击的多决策树的构造。基于针对 RFID 的攻击分为四类,所以本文构造出四个决策树,如图 4 所示。通过不同类型的攻击训练集,生成四个决策树,构造基于多决策树的入侵检测系统。其中每个决策树能够对所对应的攻击进行入侵检测,并且通过决策树之间的分析,可以得到不同攻击之间的关系。基于该异常检测模式,可以实现对攻击的识别与攻击趋势的分析,实现攻击预警。同时,系统采用基于异常检测模式的多决策树,具备实现对具有相似特征的未知攻击进行检测的能力。

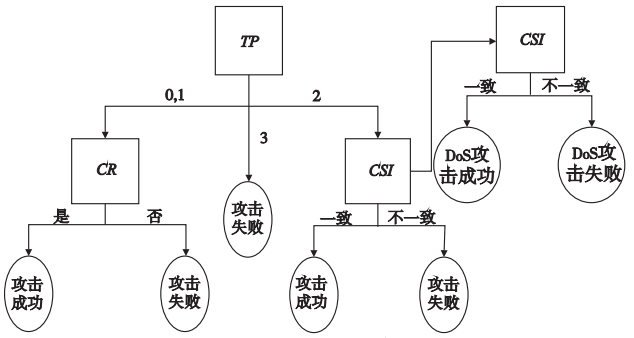

图3　跟踪攻击与DoS攻击多决策树

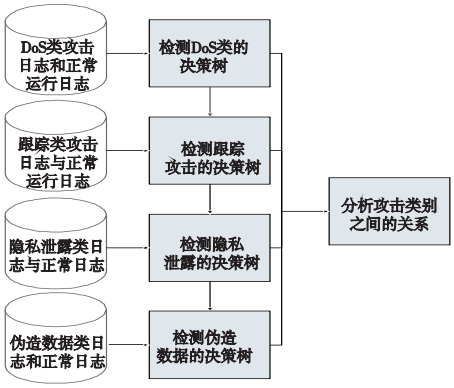

图4　多决策树入侵检测系统

4　性能分析

本文提出的基于多决策树的入侵检测参考了基于决策树的传统网络入侵检测系统的原理,并且结合了 RFID 自身的特点,决策树本身就是一个准确性较高的分类器[6,7],从而使得基于多决策树的入侵检测系统具有可扩展性和高效性。

4.1　可扩展性

由于 RFID 协议与传统的网络协议最大的区别在于 RFID 协议比较简单,所以针对 RFID 系统的攻击的不可知性远远大于传统的网络攻击。由于构造决策树主要取决于训练集和决策树自身的属性选取,所以基于单决策树的入侵检测系统最大的问题在当出现新的攻击时,单决策树的扩展性不强。本文构造多决策树,当出现不曾出现的攻击时,只用构造针对新的攻击的决策树,不用改变原有决策树的结构,从而大大提升了可扩展性。

4.2　高效性

当构造决策树时,首先必须要根据 RFID 系统的日志生成训练集。所以 RFID 的适应性是与训练集密不可分的,而训练集是由系统日志所生成,构造出的多决策树本身就是针对具体系统的,所以单个决策树自身的有效性是较高的。由于决策树和决策树之间存在链接,所以可以通过链接准确发现隐藏在系统的还未被人发现的安全隐患。

5　结束语

在基于多决策树的入侵检测模型中,系统通过对 RFID 系统的入侵行为日志和系统正常行为日志进行分析,从中得到每个攻击类的训练集和正常行为训练集。将每类攻击类的训练集和系统正常行为的训练集合并,生成四个新的训练集,每个训练集中既包含同一类的攻击训练集,也包含正常行为的训练集。通过决策树生成算法,就会生成四个决策树。每个决策树都可以区分一个系统行为是否属于所代表的攻击。通过决策树之间的连接,可以判断攻击之间的联系,从而得出对每个攻击后果的全面认识。

参考文献:

[1] SARAC A, ABSI N, DAUZÈRE-PÈRÈS S. A literature review on the impact of RFID technologies on supply chain management[J]. International Journal of Production Economics, 2010,128(1): 77-95.

[2] AHSON S A, ILYAS M. RFID handbook: applications, technology, security, and privacy[M]. [S. l.]:CRC Press, 2010.

[3] 刘欣. RFID 应用现状及面临的问题[J]. 金卡工程, 2004,8(12): 53-58.

[4] 张有光, 杜万, 张秀春, 等. 全球三大 RFID 标准体系比较分析[J]. 中国标准化, 2006(3): 61-63.

[5] AVOIN G. Cryptography in radio frequency identification and fair exchange protocols[D]. [S. l.]:Institut de systemes de communication Université de Caen, 2005.

[6] Van DEURSEN T, RADOMIROVIC S. Attacks on RFID protocols [EB/OL]. (2009-08-06). http://eprint. iacr. org/2008/310. pdf.

[7] JUELS A. RFID security and privacy: a research survey[J]. IEEE Journal on Selected Areas in Communications, 2006,24(2): 381-394.

[8] QUINLAN J R. Simplifying decision trees[J]. International Journal of Man-Machine Studies, 1987,27(3): 221-234.

[9] QUINLAN J R. Induction of decision trees[J]. Machine learning, 1986,1(1): 81-106.

[10] HAN Jia-zhi, KAMBER M, PEI Jian. Data mining: concepts and techniques[M]. 3rd ed. [S. l.]:Morgan Kaufmann, 2006.

[11] PAISE R I, VAUDENAY S. Mutual authentication in RFID: security and privacy[C]//Proc of ACM Symposium on Information, Computer and Communications Security. New York: ACM Press, 2008: 292-299.

[12] SUN Da-zhi, ZHONG Ji-dong. A hash-based RFID security protocol for strong privacy protection[J]. IEEE Trans on Consumer Electronics, 2012,58(4): 1246-1252.

[13] CHO J S, YEO S S, KIM S K. Securing against brute-force attack: a hash-based RFID mutual authentication protocol using a secret value [J]. Computer Communications, 2011,34(3): 391-397.

[14] HANATANI I Y, OHKUBO M, MATSUO S, *et al*. A study on computational formal verification for practical cryptographic protocol: the case of synchronous RFID authentication[C]//Proc of International Conference on Financial Cryptography and Data Security. Berlin: Springer-Verlag, 2012: 70-87.

[15] OSAKA K, TAKAGI T, YAMAZKI K, *et al*. An efficient and secure RFID security method with ownership transfer[C]//Proc of IEEE International Conference on Computational Intelligence and Security. 2006: 1090-1095.

[16] VERDULT R, GANS G K, GARCIA F D. A toolbox for RFID protocol analysis[C]//Proc of the 4th International EURASIP Workshop on RFID Technology. Washington DC: IEEE Computer Society, 2012: 27-34.

[17] AVOINE G, OECHSLIN P. RFID traceability: a multilayer problem [C]//Proc of the 9th International Conference on Financial Cryptography and Data Security. Berlin: Springer, 2005: 125-140.