

基于二维条码的信息隐藏技术^{*}

牛夏牧, 黄文军, 吴 迪, 张 慧

(哈尔滨工业大学信息对抗技术研究所, 黑龙江 哈尔滨 150001)

摘 要: 提出一种基于二维条码进行信息隐藏的技术。利用变形技术对二维条码 PDF417 符号字符中的各组成单元宽度加以适量的变动, 采用误差累积的方式实现隐藏信息的嵌入和提取。实验结果和安全性分析表明, 该方法可以有效的抵御打印扫描攻击, 具有安全性高、隐藏信息容量大、信息提取率高等优点。

关键词: 信息隐藏; 二维条码; 打印扫描攻击

中图分类号: TN918 **文献标识码:** A **文章编号:** 0529-6579 (2004) S2-0021-05

信息隐藏技术是信息安全研究领域最新热点之一, 其驱动力在于信息时代的两大政策问题——版权保护和状态监视。因此, 信息隐藏技术不仅关系到国计民生, 而且涉及到国家安全, 在军事和国防方面具有重大的意义。

本文提出了一种基于二维条码的信息隐藏方法, 该方法利用变形技术对二维条码符号字符的组成单元的宽度进行适量的变动以实现隐藏信息的嵌入。由于二维条码使用广泛, 体积小、容量大而且与人的视觉系统无关等特点, 因此利用二维条码作为信息隐藏的载体具有携带方便、安全性高、隐藏信息量大以及信息提取率高等优点。

利用二维条码进行隐藏通信的难点在于隐藏信息对于打印扫描攻击的鲁棒性问题。本文针对这个问题, 在 PDF417 二维条码基础上进行了仿真试验, 试验的结果证实本文提出的方法具有可行性。

1 PDF417 条码及其符号技术要求

PDF417 条码属于行排式二维条码 (见图 1)。因为组成条码的每一个符号字符都是由 4 个条和 4 个空共 17 个模块构成 (见图 2), 所以称为 PDF417 条码^[1, 2]。

PDF417 是一种多层、可变长度、具有高容量和纠错能力的二维条码。每一个 PDF417 符号可以表示 1100 字节、或 1800 个 ASCII 字符或 2700 个数字的信息。

如图 1 所示, 每行构成如下: a 为左空白区; b 为起始符; c 为左行指示符号字符; d 为 1~30 个

数据符号字符; e 为右行指示符号字符; f 为停止符; g 为右空白区。

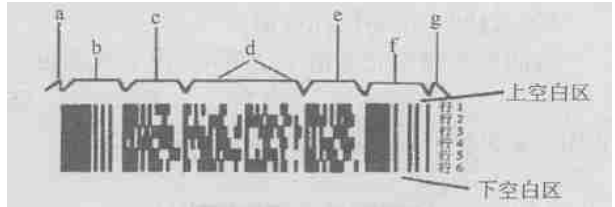


图 1 417 条码符号的结构
Fig 1 PDF417 barcode structure

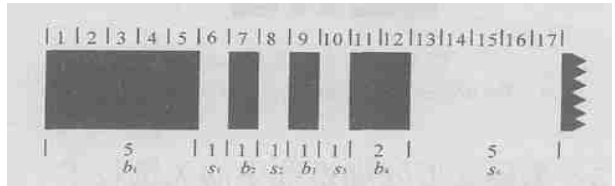


图 2 PDF417 符号字符
Fig 2 PDF417 barcode symbol character

PDF417 条码对为了提高识读率, 对条码符号的技术参数有严格的要求。因此, 条码的变形必须限制在允许的范围内, 否则将会影响到条码的正确识读。PDF417 条码的允许误差包括两个方面^[3]:

(1) 符号字符的允许误差: 在 417 符号中, 每一符号字符的误差包括符号字符允许误差 (Δp)、条或空的允许误差 (Δb)、边缘到相似边缘的允许误差 (Δe)、单元的高度允许误差 (Δh), 参见图 3。 Δp 、 Δb 、 Δe 、 Δh 分别指 p 、 b 、 e 、 h 值的允许误差, 由下式确定:

* 收稿日期: 2004-09-11

基金项目: 国家自然科学基金资助项目 (60372052); 全国博士学位论文作者专项资金资助项目 (FANEDD-200238); 黑龙江省杰出青年科学基金资助项目; 哈尔滨工业大学跨学科研究基金资助项目 (HIT MD-2002. 11)

作者简介: 牛夏牧 (1961 年生), 男, 教授, 博士生导师; E-mail: xianu niu@hit.edu.cn

$$\Delta b = \pm (0.40x - 0.0127) \tag{1}$$
$$\Delta e = \pm 0.20x \tag{2}$$
$$\Delta p = \pm 0.20x \tag{3}$$
$$\Delta h = \pm 0.20x \tag{4}$$

式中， x 为符号的模块宽，单位为 mm。

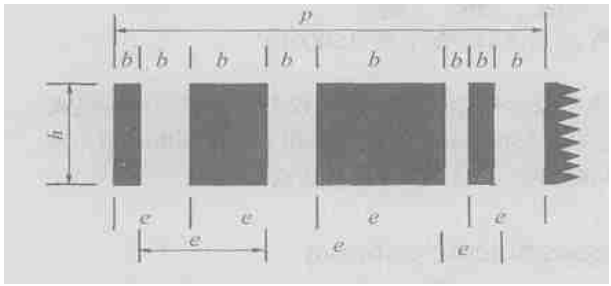


图 3 PDF417 条码允许误差
Fig. 3 PDF417 barcode tolerance

(2) 行间水平错位允许值。

行间水平错位允许值 a (图 4) 由下式确定：

$$a = \pm 0.12x \tag{5}$$

式中， x 为符号的模块宽，单位为 mm。

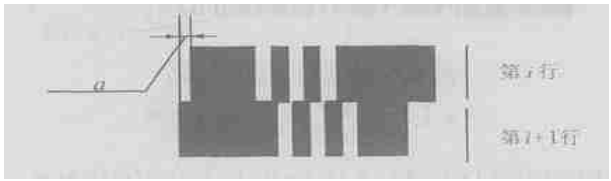


图 4 行间水平错位允许值

Fig. 4 The tolerance of horizontal shift between the line space

2 条码变形处理系统中的关键技术

本文利用变形技术在条码中实现信息的隐藏^[4]，对条码符号字符的组成单元宽度引入适量的误差，利用误差累积实现隐藏信息的嵌入与提取^[5]。条码变形处理系统包括隐藏信息的嵌入、检测和提取、对隐藏信息的攻击和隐藏信息数量等。

2.1 信息的嵌入

PDF417 条码中，隐藏信息选择嵌入的区域从每行的左行指示符号字符到右行指示符号字符，即图 1 中 $c-e$ 的区域。起始符与终止符不做任何的改变，这种做法可以解决一般变形技术在解码时需要原始图像信息的缺陷，即实现盲提取。

2.1.1 嵌入信息“1” 如果嵌入的隐藏信息为“1”，那么，对于隐藏信息嵌入区域中的每一个符号字符，将构成其的四个单元条中增加适量的误差 Δw ，而边缘到相似边缘的值保持不变（即图 3 中的 e 值保持不变），这时，与每个条相邻的空将自动减小 Δw 的宽度。由于条与空的引入误差值相应

的抵消，因此，对于条码符号字符的宽度并没有改变（即图 3 中的 p 值）。

对于一个 PDF417 符号字符而言，由于在它的每个构成单元条中增加了 Δw 的误差值，因此，4 个条的总误差：

$$T_b = (w'_{b_1} - w_{b_1}) + (w'_{b_2} - w_{b_2}) + (w'_{b_3} - w_{b_3}) + (w'_{b_4} - w_{b_4}) = 4\Delta w \tag{6}$$

相应的，4 个空的总误差：

$$T_s = (w'_{s_1} - w_{s_1}) + (w'_{s_1} - w_{s_1}) + (w'_{s_1} - w_{s_1}) + (w'_{s_1} - w_{s_1}) = -4\Delta w \tag{7}$$

因此，一个符号字符的误差为：

$$T = T_b - T_s = 4\Delta w - (-4\Delta w) = 8\Delta w \tag{8}$$

如果每 n 个符号字符嵌入一位隐藏信息，那么， n 个符号字符的误差

$$T_n = nT = 8n\Delta w \tag{9}$$

这样，可以把加入的误差值 $8n\Delta w$ 看作嵌入隐藏信息“1”。

2.1.2 嵌入信息“0” 如果嵌入的信息为“0”，对于隐藏信息嵌入区域中的每一个符号字符，将构成其的 4 个单元条中减少适量的误差 Δw 。边缘到相似边缘的值保持不变（即图 3 中的 e 值保持不变），同时，与每个条相邻的空将相应的增加 Δw 的宽度。由于条与空的引入误差值相应的抵消，因此，对于条码符号字符的宽度并没有改变（即图 3 中的 p 值）。

对于一个 PDF417 符号字符而言，由于在它的每个构成单元条中减少了 Δw 的误差值，因此，4 个条的总误差：

$$T_b = (w'_{b_1} - w_{b_1}) + (w'_{b_2} - w_{b_2}) + (w'_{b_3} - w_{b_3}) + (w'_{b_4} - w_{b_4}) = -4\Delta w \tag{10}$$

相应的，4 个空的总误差：

$$T_s = (w'_{s_1} - w_{s_1}) + (w'_{s_1} - w_{s_1}) + (w'_{s_1} - w_{s_1}) + (w'_{s_1} - w_{s_1}) = 4\Delta w \tag{11}$$

因此，一个符号字符的误差为：

$$T = T_b - T_s = -4\Delta w - 4\Delta w = -8\Delta w \tag{12}$$

如果每 n 个符号字符嵌入一位隐藏信息，那么， n 个符号字符的误差

$$T_n = nT = -8n\Delta w \tag{13}$$

这样，可以把加入的误差值 $-8n\Delta w$ 看作嵌入信息“0”。

2.1.3 嵌入算法分析 从上面的描述可以看出，在信息嵌入算法中引入的误差基本上不影响条码的尺寸规范。因为构成每一个条码符号的条与空的误差相互抵消，因此，符号字符允许误差 (Δp)、边缘

到相似边缘的允许误差(Δe)、单元的高度允许误差(Δh)以及行间水平错位允许值 a 等参数并没有受到影响。唯一受到影响的参数是条或空的允许误差(Δb)。只要单元条中引入的误差值 $\Delta w < \Delta b$,那么,条码变形产生的误差将符合PDF417条码符号的技术要求。

2.2 信息的检测和提取

隐藏的检测过程不需要原始信息参与,属于盲检测过程。隐藏信息的提取按照下列的步骤进行:

(1) 计算变形后条码中起始符和终止符模块的平均宽度 x_0 ,将其作为条码模块的标准宽度。

$$x_0 = \frac{1}{u} \left(\sum_{i=1}^u l_{si} + l_{pi} \right) \setminus 35 \tag{14}$$

其中, l_{si} 为第 i 行起始符的宽度, l_{pi} 为第 i 行终止符的宽度, u 为条码符号的行数。35为起始符的17个模块与终止符的18个模块的和。

(2) 根据 x_0 计算条码符号中每一个单元的宽度 w ,然后与单元的实际宽度 w' 做差,得到每一个条码符号中单元的误差值 Δw 。

$$w = w' - w = kx - kx_0 \tag{15}$$

其中 k 为组成单元的模块个数, x 为该单元模块的实际宽度。

(3) 假设每 n 个条码字符嵌入一位隐藏信息。考虑到打印扫描以及其他因素对条码的影响,因此,对条码中隐藏信息的提取应该设置一个阈值 s ($0 < s < 8nk\Delta x$),如果检测出的误差值 $T_n > s$,认为在这 n 个条码字符中嵌入的信息为“1”;如果检测出的误差值 $T_n < -s$,认为在这 n 个条码字符中嵌入的信息为“0”。

2.3 攻击

对于二维条码变形技术来说,主要的攻击来自打印和扫描,经过打印和扫描后的条码图像会产生轻微的变形,这种变形不但会影响到隐藏信息的检测和提取,同时还会影响到条码的正常识读。

二维条码变形技术与图像变形技术^[6,7]和文档的变形技术不同^[8-12],它与条码的识读有着密切的关系,而与人的视觉系统无关,因此隐藏效果较好;同时,也导致了对二维条码隐藏信息抗攻击鲁棒性的描述不能采用以往的评价标准^[13,14],必须寻找其它适合的评价指标。本文根据条码的识读,初步提出了4种评价指标,分别是抗噪率 γ 、最小抗噪率 $\gamma_{\min}$ 、隐藏信息对条码识读的影响度 ξ 和稳定度 η 。

(1) 抗噪率 γ :表示二维条码经过变形技术处理以后,隐藏信息嵌入区域中,嵌入信息的符号字符误差平均值 T_w 与原始条码符号字符误差平均值

T_o 之间的关系。定义为

$$\gamma = 20 \lg \frac{T_w}{T_o} \tag{16}$$

(2) 最小抗噪率 $\gamma_{\min}$:表示二维条码经过变形技术处理以后,其嵌入信息的符号字符中最小误差值 $T_{w\min}$ 与原始条码符号字符最大误差值 $T_{o\max}$ 之间的关系。定义为:

$$\gamma_{\min} = 20 \lg \frac{T_{w\min}}{T_{o\max}} \tag{17}$$

(3) 隐藏信息对条码识读的影响度 ξ :表示条码隐藏信息载荷与拒读错误数 e 的关系。定义为:

$$\xi = \frac{e}{w} \times 100\% \tag{18}$$

(4) 隐藏信息的稳定度表示变形的二维条码在经过打印扫描攻击后,检测出变形范围的稳定度。定义为:

$$\eta = \frac{T_{w\max}}{T_{w\max} - T_{w\min}} \tag{19}$$

2.4 隐藏信息的数量

二维条码可隐藏的信息量是非常大的。前面提到,隐藏信息选择嵌入的区域从每行的左行指示符号字符到右行指示符号字符,如果每一个符号字符嵌入一位隐藏信息,最大嵌入容量 w 为

$$w = u \times (v - 2) \tag{20}$$

这对于具有几十行、几十列,而尺寸仅为数cm的条码来说,其嵌入容量是惊人的。如果采用高精度的打印和扫描设备,甚至可以在符号字符中每一个组成单元条或空嵌入一位隐藏信息,这时条码隐藏信息容量将以倍数增长。

2.5 安全性

利用二维条码进行信息隐藏具有较高的安全性:①条码的应用非常广泛,以条码为载体不易引起察觉;②基于二维条码的信息隐藏与人的视觉系统无关,人眼无法分辨;③条码隐藏信息容量大,提取率高,而尺寸较小,便于携带。

3 仿真试验

仿真实验主要是针对条码的打印扫描攻击,根据前面所描述的条码隐藏信息抗攻击鲁棒性的技术指标对实验结果进行分析。

实验采用3行7列的PDF417条码进行信息的嵌入与提取,条码的编码信息为“PDF417”,纠错级别为1级,宽高比1:3,表1表示“PDF417”的417条码的条、空组合形式。条码的模块宽度设定为1mm,对隐藏信息嵌入区域中每一个符号字符

组成单元引入的误差为 0.25 mm。在隐藏信息嵌入区域中的每一个符号字符嵌入一位隐藏信息，隐藏信息总数为 15，信息选择“111110000011111”。图 5 为未经过打印扫描的条码图形，其中图 5a 为原始条码，图 5b 为变形条码。可以看到由于加入的误差值较小，人的肉眼很难察觉出来。

变形后的条码和原始条码分别用 200 dpi 和 300 dpi 的分辨率进行扫描。扫描后原始条码图像与变形条码图像中各种实验参数及结果见表 2。表 3 列出了不同扫描分辨率下，条码隐藏信息抗攻击鲁棒性的各项指标的比较。

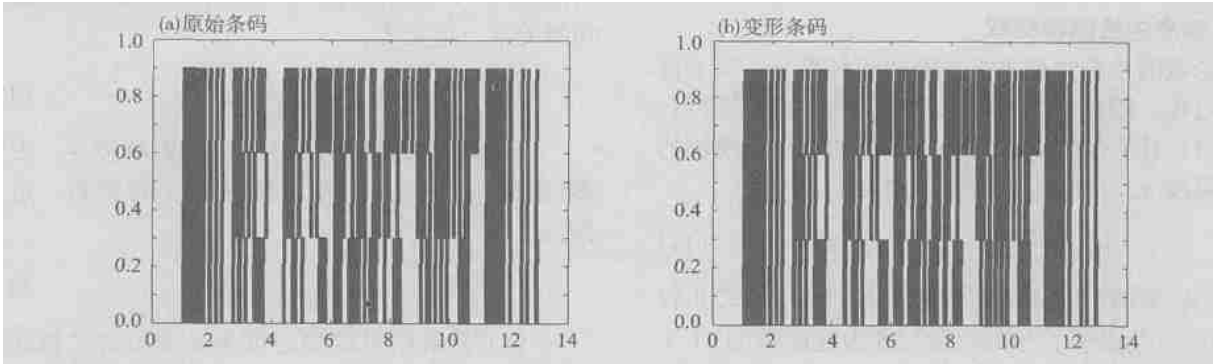


图 5 未经扫描的 PDF417 条码
Fig. 5 The PDF417 barcode before scan

表 1 表示“DF417”的 417 条码的字、空组合形式
Tab. 1 The format of bar and space of PDF417

行数	bsbsbsbs	bbsbsbsbs	bsbsbsbs	bsbsbsbsbs	bsbsbsbs	bsbsbsbsbs	bsbsbsbsbs
第 1 行	81111113	31111136	51111251	31312223	51211142	51111152	711311121
第 2 行	81111113	41111315	41131115	42113231	22163111	51111125	711311121
第 3 行	81111113	11111246	12211532	32411132	12361121	11111345	711311121

表 2 不同扫描分辨率下的条码实验参数
Tab. 2 The experimental results in the different scan resolution

扫描分辨率 dpi	条码	T_{max} mm	T_{min} mm	T mm	T_{bls} mm	二值化阈值 mm	s mm	隐藏信息 提取率/%	拒读错 误数 e
200	变形条码	2.625	1.125	1.830	0.387	180	1.125	100	1
	原始条码	0.500	0	0.234	0.387	180	—	—	0
300	变形条码	2.500	1.580	1.882	0.387	180	1.580	100	1
	原始条码	0.250	0.080	0.157	0.387	180	—	—	0

表 3 不同扫描分辨率下隐藏信息
抗攻击鲁棒性评价指标比较

Tab. 3 The benchmarking results in the different scan resolution

扫描分辨率 dpi	γ %	γ_{min} %	ξ %	η
200	17.8675	7.044	6.67	1.75
300	21.579	16.014	6.67	2.72

4 结 论

从以上对实验结果的分析得到以下结论：①本

文提出的隐藏算法能够有效的抵抗打印和扫描攻击；②信息隐藏的效果较好，肉眼难以分辨；③经过较高扫描分辨率扫描后对隐藏信息提取的效果要优于低分辨率。
信息隐藏是目前信息时代研究领域的热点，在国家安全方面具有重要的意义。本文使用变形技术对二维条码进行处理以实现信隐藏通信。
从实验结果和安全分析可以看出，本文提出的方法可以有效的抵御打印扫描攻击、隐藏信息容量大，隐蔽效果好，提取率高、使用方便等优点。

参考文献:

[1] 陈平,周克俭,张丙齐. 二维条形码(PDF417)及其应用[J] . 计算机应用, 1999 1; 15— 17.

[2] 梁凤梅. PDF417 条码及其识读[J] . 电脑开发与应用. 2002, 15(6); 17— 19.

[3] GB/T 17172— 1997. 四一七条码[S] . 国家技术监督局, 1997.

[4] STEFAN K F, PETITCOLAS A P. 信息隐藏技术—— 隐写术语数字水印[M] . 吴秋新, 等译. 北京: 人民邮电出版社, 2001; 49— 53.

[5] 牛夏牧, 黄文军, 张慧. 一种堆栈式条码的防伪方法[P] . 200410013532 5, 2004.

[6] STANDFORD M T, VRADLEY J N, HANDEL T G. Data embedding method[C] . Proceedings of the SPIE 2615, Integration Issues in Large Commercial Media Delivery Systems, 1996; 561— 572.

[7] STANDFORD M T, HANDEL T G, ETTINGER J M. Data embedding in degenerate host[M] . Technical Report LA— 95— 4446UR, Los Alamos National Laboratory, 1996.

[8] BRASSIL J, et al. Document marking and identification using both line and word shifting [C] . Proceedings of INFOCOM '95, 1995; 853— 860.

[9] MACENCHUK N F. Electronic document distribution[C] . At & T Technical J, 1994; 73— 80.

[10] LOW S H, MAXEMCHUK N F, LAPONE A M. Document marking and identification for copyright protection using centroid detection [J] . IEEE Transactions on Communications, 1998, 46(3); 372— 383.

[11] LOW S H, MAXEMCHUK N F. Performace comparison of two text marking methods[J] . IEEE J Selected Areas in Communications, 1998, 16(4); 561— 572.

[12] BRASSIL J, MAXEMCHUK N F, O'GORMAN L. Electronic marking and identification techniques to discourage document coping[C] . Proceedings of Inforcom '94, 1994; 1278— 1287.

[13] JOHNSON N F, JAJODIA S. Steganalysis of images created using current steganography software[C] . Proceedings of the Second International Workshop on Information Hiding, 1998, 1525; 273— 289.

[14] PETICOLAS F A P, ANDERSON R, KUHN M. Attacks on copyright marking systems[C] . Proceedings of the Second International Hiding, 1998, 1525; 218— 238.

Information Hiding Technique Based on 2D Barcode

NIU Xia mu, HUANG Wen jun, WU Di, ZHANG Hui

(Information Countermeasure Technique Research Institute, Ha' erbin Insitute of Technology, Ha' erbin 150001, China)

Abstract: This paper proposed a information heding technique 2D barcode. By modifying the width of each uniof PDF417 barcode, information is hidden into the barcode. The experimental results show that the algorith features robustness, security, and high capacity of information hiding etc.

Key words: information hiding; 2D barcode; attacks of print scan