基于搜索机制混沌加密算法的密文特性分析

葛 辛　　刘粉林　　芦 斌　　陈 娟
(解放军信息工程大学信息工程学院　郑州　450002)

摘　要：针对Baptista提出的混沌加密算法及其改进算法中的密文过长及 0, 1 比例不均匀的问题，该文通过计算密文的信息熵，给出了密文明文比例期望的下界；然后运用文中所提出的 N- 截断等长编码和最优 N- 截断等长编码等概念描述了编码方式与密文明文比例之间的关系，给出了 0, 1 比例的近似计算公式。算例分析表明最优 N- 截断等长编码可有效缩短密文长度，提高 0, 1 比例的均匀度。

关键词：混沌加密；信息熵；最优 N- 截断等长编码
中图分类号：TP309　　**文献标识码**：A　　**文章编号**：1009-5896(2008)07-1625-05

Analysis for the Cipherext Characteristic of Based on Search Mechanism Chaotic Cryptosystem

Ge Xin　　Liu Fen-lin　　Lu Bin　　Chen Juan
(*Information Engineering Institute*, *the PLA Information Engineering University*, *Zhengzhou* 450002, *China*)

Abstract: In order to solve the two major drawbacks of Baptista's chaotic cryptosystem and its modified versions——excessive length of ciphertext and unbalance ratio of bit 0 to bit 1, this paper gives the lower bound of the expectation of the cipher-to-plaintext ratio by calculating the entropy of the ciphertext. Then exploiting N- truncated equal length coding, the optimal N- truncated equal length coding and so on, this paper describes the relationship between coding and cipher-to-plaintext ratio, and proposes the approximate formula to work out the ratio of bit 0 to bit 1. Numerical paradigms show that the optimal N- truncated equal length coding can efficiently shorten the length of ciphertext and obtain more uniform ratio of bit 0 to bit 1.

Key words: Chaos cryptosystem; Information entropy; Optimal N- truncated equal length coding

1　引言

随着混沌密码研究的深入，相继出现了许多基于混沌的加密算法，其中 Baptista 等人 1998 年提出的基于搜索机制的混沌加密算法[1]就是在这方面的积极尝试。

Baptista 算法自提出以来引起了许多学者的广泛关注[2−13]，一些学者相继提出了一系列改进算法[2−10]。针对密文分布不均匀问题，W K Wong 于 2001 年提出在加密每组明文之前预迭代混沌系统 K 次(K 为随机数)，降低了加密速度[2]。李树钧等人尝试对密文进一步加密以改善 Baptista 算法中密文分布，但以增加算法实现复杂度为代价，加密速度较慢[7, 8]。K W Wong 等人从不同角度对 Baptista 算法进行了改进。首先，提出动态更新 Baptista 算法中的双射(文中称为查找表)，并去掉原算法中的附加参数以获得更快的加密速度[4]；其次在文献[5]中进一步提出了增强的动态查找表技术，同时引入一个会话密钥以增强安全性，并使用每个明文字符在动态查找表中的索引值作为密文，有效地降低了密文长度；文献[6]据加密每组明文对应的查找表不同，将查找表作为消息认证码，并引入加密参数，动态决定每组明文的加密次数，以提高算法的安全性；而文献[9]进一步改进了查找表的更新方式，把查找表的大小扩充到 2^{n+1} (n 为明文分组的长度)。针对 Baptista 算法的安全性问题，文献[3]建议使用多个耦合混沌映射网络中产生的交替混沌；文献[10]提出使混沌参数与明文相关、动态更新混沌参数来增强原加密算法的安全性。与此同时，一些学者也对此类算法的安全性提出了质疑，相关结论见文献[11−13]。

上述改进算法主要从加密过程入手改进 Baptista 算法存在的缺陷，同时也尝试利用不同的密文编码方式来改进密文长度过长及密文中 0, 1 比例不均匀的问题[1, 2, 4, 9, 10]。然而，截至目前，未见相关文献深入探讨编码方式对此类问题的影响。那么密文可缩短到的最短长度是多少？若存在最短长度，那么是否存在某种编码方式可达到或接近此最短密文长度？通过何种编码方式能使 0, 1 比例均匀或接近于均匀？这是一系列值得探讨的问题。

本文正是针对这些问题，以信息论和编码理论为基础，使用密文明文比例来刻画密文长度，利用信息熵给出了密文明文比例期望的下界；进而提出了 N- 截断等长编码(N-Truncated Equal Length Coding, N-TELC)及最优 N- 截断

2006-12-18 收到，2007-06-08 改回
国家自然科学基金(60374004)，河南省杰出青年基金(0412000200)，国家 863 项目(2006AA01Z409)和河南省 06 科技攻关项目资助课题

等长编码(Optimal N-Truncated Equal Length Coding, N-OTELC)，并描述了编码方式与密文长度之间的关系，给出了 0, 1 比例的近似计算方法(使用 bit 0 在密文中所占的比例来衡量[14])。算例分析表明采用 N-OTELC 可有效缩短密文长度，提高 0, 1 比例的均匀度。

2 Baptista 及其改进算法中密文的信息熵

2.1 Baptista 及其改进算法的基本思想

Baptista 及其改进算法的基本思想可以用下面的七元组 $(n,F,T,\varphi,\nu,\mathrm{Enc},\mathrm{Dec})$ 来描述，其中：

(1) n 表示把待加密的明文 m 按 n bit 分组，分组后的明文可表示为 $m_1,m_2,\cdots$，其中第 i 组明文 $m_i\in\mathcal{M}=\{0,1,2,\cdots,2^n-1\}$，$\mathcal{M}$ 即为明文空间。

(2) F 为一合适的混沌系统，其初始状态 x_0 和混沌控制参数 b 作为密钥。

(3) T 为 F 状态空间的子空间(可以为一部分或者全部)，将 T 均匀划分为 2^n 个部分 T_i，$i=1,2,3\cdots,2^n$。记 $\mathcal{T}=\{T_1,T_2,\cdots,T_{2^n}\}$。

(4) φ 是双射，且 $\varphi\in\mathcal{A}=\{f:\mathcal{M}\to\mathcal{T}\}$。

(5) ν 是自然数集 $\mathcal{N}\to\{0,1\}$ 上的映射，其作用为在 $\mathcal{N}$ 中选择合适的子集作为密文空间。记 $C=\{k\in\mathcal{N}\mid\nu(k)=1\}$，$C$ 为密文空间。

(6) Enc 为加密方案，其描述如下：加密前置 $x_0^{(1)}\leftarrow x_0$；加密第 i 组明文 m_i：以 $x_0^{(i)}$ 为初始状态，对混沌状态 F 进行迭代。若对于第 k 次迭代，满足 $F^k(x_0^{(i)})\in\varphi(m_i)$ 且 $k\in C$，则终止迭代。此时，令 $C_i=k$ 为 m_i 的密文，并令 $x_0^{(i+1)}\leftarrow F(x_0^{(i)})$。

(7) Dec 为解密方案，其描述如下：解密前置 $x_0^{(1)}\leftarrow x_0$；解密第 i 组密文 C_i：以 $x_0^{(i)}$ 为初始状态，对混沌系统 F 迭代 C_i 次，若 $F^{C_i}(x_0^{(i)})\in\varphi(m_i)$，则 m_i 为 C_i 对应的明文；并令 $x_0^{(i+1)}\leftarrow F^{C_i}(x_0^{(i)})$。

不同的算法中，ν 的取值不同。但在最新的改进算法中[10]，ν 在 $\mathcal{N}$ 上取值均为 1，也就是说，ν 选取的密文空间 C 为 $\mathcal{N}$。本文下面均设密文空间 C 为 $\mathcal{N}$ 进行讨论。

2.2 Baptista 及其改进算法中密文信息熵的计算

本节讨论 2.1 节描述的加密算法中密文的信息熵问题。为此据混沌系统相关性质假设：在理想状态下，混沌系统每次迭代后的状态落入各个区间的概率相同[15, 16]。

本文讨论明文分组的一般形式，即明文采用 n bit 分组，把混沌系统的状态空间均匀划分为 2^n 个区间 $T_1,T_2,\cdots,T_{2^n}$，故加密每组明文后的密文 C_i 可以看成一个随机变量，该随机变量服从几何分布，即

$$P\{C_i=\mu\}=p(1-p)^{\mu-1} \tag{1}$$

其中 $p=1/2^n$，$\mu=1,2,\cdots$。则密文 C_i 的信息熵为(密文将使用 bit 表示，故对数函数的底取 2)

$$H(C_i)=-\left[\log_2 p+(1-p)\log_2(1-p)/p\right] \tag{2}$$

即：加密一组明文后的密文(后称一组密文)所包含信息量的统计平均值为 $H(C_i)$。因此密文明文比例期望的下界为

$$R_m(n)=1-(1-p)\log_2(1-p)/(np) \tag{3}$$

根据以上事实，存在如下命题：

命题 1 在 2.1 节所描述的方案中(明文采用 n bit 分组)，若混沌系统状态落入每个小区间的概率相等，则 $R_m(n)$ 关于 n 单调递减，且 $\lim\limits_{n->\infty}R_m(n)=1$。(证明略)

根据式(3)，表 1 给出了明文分别采用1,2,⋯,16 bit 分组时的 $R_m(n)$ 的值，例如，第 2 行第 3 列的1.622556 意味着：如果明文采用 2bit 分组，则密文明文比例的期望不可能小于 1.622556。

3 *N*-OTELC

上节利用信息熵给出了密文明文比例期望的下界。那么如何编码可使编码后密文明文比例的期望尽量接近其下界？

3.1 *N*-OTELC

定义 1 把正整数 A 表示成“$\underbrace{N-1,N-1,\cdots,N-1}_{(k-1)个},w$”的形式，即使用“$\underbrace{N-1,N-1,\cdots,N-1}_{(k-1)个},w$”这 k 个整数表示 A。其中，N 为正整数且 $N\ge 2$，$k=\left\lfloor\dfrac{A-1}{N-1}\right\rfloor+1$，$w=(A-1)\bmod(N-1)$，则称该过程为 N-截断，N 为截断位置。

定义 2 对正整数 A 做 N- 截断后，A 可以使用有限个码元 $\{0,1,\cdots,N-1\}$ 表示，如果每个码元使用 $\lceil\log_2 N\rceil$ bit 编码(其中 N 为截断位置，$N=2,3,\cdots$)，则称这种编码方式为 N-TELC。

定义 3 在 2.1 节描述的加密方案中，明文采用 n bit 分组，编码后的密文最短的 N-TELC 称为 n bit 的 N-OTELC。

由 N-TELC 定义易知，使用 N-TELC 对 A 编码后为 $\underbrace{N-1,N-1,\cdots,N-1}_{(k-1)个},w$，其长度(单位 bit)为

$$L(A)=k\lceil\log_2 N\rceil \tag{4}$$

表 1 密文明文比例期望的下界

n(bit)	1	2	3	4	5	6	7	8
$R_m(n)$	2.000000	1.622556	1.449505	1.349160	1.283983	1.238561	1.205292	1.179984
n(bit)	9	10	11	12	13	14	15	16
$R_m(n)$	1.160143	1.144199	1.131122	1.120210	1.110970	1.103047	1.096178	1.090168

则编码 C_i 后的码长 $L(C_i)$ 为 $\lceil \log_2 N \rceil$ 的整数倍， $L(C_i)=k\lceil \log_2 N \rceil$， $k=1,2,\cdots$ 的概率为

$$P\{\ L(C_i)=k\lceil \log_2 N \rceil\ \}=(1-p)^{(k-1)(N-1)}[1-(1-p)^{N-1}] \quad (5)$$

因此 $L(C_i)$ 的期望(平均码长)为

$$E[L(C_i)]=\frac{\lceil \log_2 N \rceil}{1-(1-p)^{N-1}} \quad (6)$$

其中 $p=1/2^n$ 。则密文明文比例的期望为

$$E_r=\frac{1}{n}\times E[L(C_i)] \quad (7)$$

据式(7)知，明文采用 n bit 分组， E_r 是截断位置 N 的函数，故可设 $h(N)=\dfrac{\lceil \log_2 N \rceil}{n\ [1-(1-p)^{N-1}]}$，则有如下命题：

命题 2　当截断位置 $N\in\{2^{m-1}+i\mid i\in 1,2,\cdots,2^{m-1},\ m\geq 1, m\in Z^+\}$ 时，密文明文比例的期望随截断位置 N 的增大而减小，且 N 取 2^m $(m\geq 1, m\in Z^+)$ 时，密文明文比例的期望最小。(证明略)

由式(7)及命题 2 知，若密文采用 N-TELC，在集合 $\{2^{m-1}+i\mid i\in 1,2,\cdots,2^{m-1}, m\geq 1, m\in Z^+\}$ 上， N 取 $2^m(m>1, m\in Z^+)$时密文明文比例的期望最小，此时密文明文比例的期望为

$$E_r=\frac{m}{n\ [1-(1-p)^{2^m-1}]} \quad (8)$$

如明文为 n bit 分组，设 $f(m)=\dfrac{m}{n\ [1-(1-p)^{2^m-1}]}$ $(p=1/2^n$， m, n 为正整数， n 为常数)，则存在如下命题：

命题 3　在 N-OTELC 下，密文明文比例的期望存在最小值为 $\min\{f(1),f(2),\cdots,f(2^n)\}$。 (证明略)

由命题 3 可知， m_0 的确定只需穷举有限个实数 $f(1), f(2),\cdots,f(2^n)$ 即可。当 $m=m_0$ 时，密文明文比例取到最小值，即编码后密文长度最短，因此 n bit 下的 N-OTELC 的截断位置 $N_0^{(n)}=2^{m_0}$ ， 即 2^{m_0}- TELC 是 n bit 下的 N-OTELC。此时的密文明文比例的期望为

$$E_r=\frac{m_0}{n\ [1-(1-p)^{2^{m_0}-1}]} \quad (9)$$

注: n 取值不同时，对应的 m_0 取值也不同，即 N-OTELC 也不同。

3.2 用 N-OTELC 编码后密文中 bit 0 所占比例

一般地，设 $g_0(x)$， $g_1(x)$ $(x\in[0,N-1])$分别表示把 x 使用二进制表示后 bit 0 , bit 1 的个数。据定义 2，对正整数 A 采用 N-TELC 编码，结果为：$\underbrace{N-1,N-1,\cdots,N-1}_{\lfloor (A-1)/(N-1)\rfloor 个},(A-1)\cdot \mathrm{mod}(N-1)$。所以整数 A 的编码结果中 bit 0 , bit 1 的个数分别为

$$G_j(A)=\left\lfloor \frac{A-1}{N-1}\right\rfloor \times g_j(N-1)+g_j((A-1)\mathrm{mod}(N-1)) \quad (10)$$

则对 C_i 进行编码后，其中 bit 0 , bit 1 个数的期望分别为

$$E_j=\lim_{s\to\infty}\sum_{A=1}^{s}[P\{C_i=A\}\times G_j(A)] \quad (11)$$

其中 $j=0,1$ 。密文中 bit 0 所占比例的期望 $R=E_0/(E_0+E_1)$。由于 E_0， E_1 计算困难，可采用如下方法近似计算：

$$R\approx R'=\frac{E_0'}{E_0'+E_1'}=\frac{\sum_{A=1}^{J}[P\{C_i=A\}\times G_0(A)]}{\sum_{A=1}^{J}[P\{C_i=A\}\times G_0(A)]+\sum_{A=1}^{J}[P\{C_i=A\}\times G_1(A)]} \quad (12)$$

其中 J 为正整数。

3.3 N-OETLC 与 N-ETLC 的比较

前文从理论上证明了，在所有 N-TELC 中，使用 N-OTELC 所得到的密文明文比例期望最小，并给出了 0, 1 比例的近似计算公式。由命题 3 可知，当明文采用 4 bit 分组时， N-OTELC 的截断位置 $N_0^{(4)}=2^5$ 。本节以明文 4 bit 分组为例说明上述问题，从密文明文比例的期望和 bit 0 所占比例的近似值两个方面比较了 N-OTELC 与其他 N-TELC 的性能。计算结果如图 1 所示。图 1(a)(根据式(7)计算)显示了截断位置与密文明文比例期望值之间的关系。直线表示的是明文采用 4 bit 分组时，根据命题 1 得到的密文明文比例期望的下界 1.349160 。由图 1(a)可知， N-TELC 后，密文明文比例的期望没有突破该下界，这验证了命题 1 的结论；当 $N\in\left\{2^{m-1}+i\mid i\in 1,2,\cdots,2^{m-1}, m\geq 1, m\in Z^+\right\}$ 时，密文明文比例的期望随着 N 的增大而减小，这验证了命题 2 的结论；在截断位置所有可能的取值中，当 $N=2^5$ 时密文明文比例的期望最小为 1.445500 ，即当 $n=4$ 时， $f(5)=\min\{f(1), f(2), f(3),\cdots,f(2^n)\}=1.445500$ ，这验证了命题 3 的结论。图 1(b)是根据式(12)计算的采用 N-TELC 后 bit 0 所占比例的近似值，其中 J 取满足 $P\{C_i=J\}<10^{-100}$ 的最小正整数。采用 2^5- TELC 时 bit 0 所占比例约为 0.51978 ，接近于 0.5 。显然， 2^5- TELC 的性能比其他截断位置 N-TELC 的性能要好。

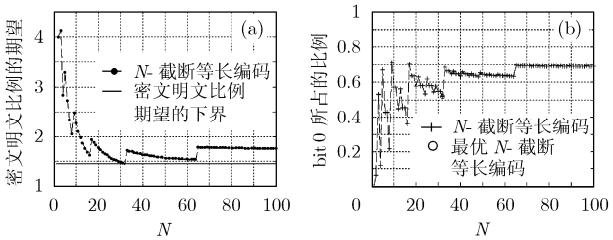

图 1 N-TELC 下密文明文比例的期望与 bit 0 所占比例

4　算例及其分析

本节实现了第 2.1 节描述的加密方案，并采用 N-OTELC 对密文进行编码，下面将详细讨论和分析实验结果。在实现中，使用的混沌系统为

$$x_{i+1}=\begin{cases}x_i/b, & x_i\leq b\\ (1-x_i)/(1-b), & x_i>b\end{cases} \quad (13)$$

表 2 *N*-OTELC 的参数表

n(bit)	1	2	3	4	5	6	7	8
$N_0^{(n)}$	2^1	2^2	2^4	2^5	2^6	2^7	2^9	2^{10}
E_r	2.000000	1.729730	1.541308	1.445490	1.387785	1.349262	1.309511	1.273229
R	0.500000	0.414062	0.545567	0.519778	0.503564	0.492540	0.584766	0.575276
n(bit)	9	10	11	12	13	14	15	16
$N_0^{(n)}$	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}	2^{18}
E_r	1.244980	1.222367	1.203857	1.188428	1.175371	1.164179	1.154478	1.145989
R	0.567552	0.561136	0.555718	0.551078	0.547059	0.543545	0.540444	0.537688

控制参数 $b = 0.67777$，初始值 $x_0 = 0.17777$。密文选择函数 ν 在 $\mathcal{N}$ 上取值全为 1，与最新的改进算法中相同[10]。在实现混沌系统时，运算中使用 IEEE 规定的 64bit 表示的双精度浮点数[17]。明文为一个从 Internet 网上下载的一个 184K 的 Mp3 文件(网址 http://site.waps.cn:82/files/20060325/16095926.mp3)。实验过程中 *N*-OTELC 的参数值见表 2。

表 2 给出了明文分别采用 1, 2, …, 16 bit 分组时，*N*-OTELC 的截断位置(基于命题 3 计算)、密文明文比例的期望 E_r (使用式(9)计算)、 bit 0 所占比例期望 R 的近似值(根据式(12)，J 取满足 $P\{C_i = J\} < 10^{-100}$ 的最小正整数计算)。图 2 给出了实际加密后密文明文比例和 bit 0 所占比例的试验值，并将他们与表 2 中的理论值进行比较。从图 2 中可以看出，除明文使用 2 bit 以下分组时，期望与试验值相差较大外，其他均吻合较好。造成这种差异的原因是，混沌系统迭代后的状态分布与假设的几何分布并不完全一致。

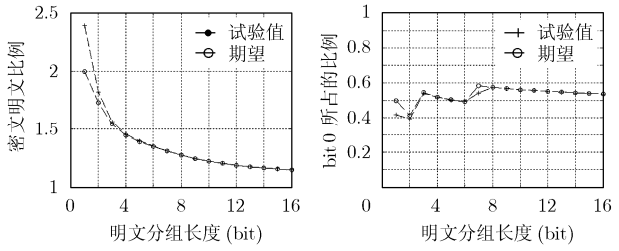

图 2 *N*-OTELC 下密文明文比例与
bit 0 所占比例试验值与理论期望对比

5 结束语

本文针对 Baptista 等人提出的基于搜索机制的混沌加密方案及其改进方案中存在的两个密文特性方面的问题——密文过长和 0, 1 比例不均匀进行了分析。给出了 0, 1 比例的近似计算方法，通过求密文的信息熵给出了密文明文比例期望的下界，然后给出改善密文特性的方案——*N*-OTELC。理论和算例分析都表明采用 *N*-OTELC 可有效缩短密文长度，提高 0, 1 比例的均匀度。但采用 *N*-OTELC 后，密文明文比例还未达到其下界，故还存在进一步缩短密文长度的可能。那么是否存在一种编码方式可以进一步缩短密文长度是值得深入探讨的问题。

参 考 文 献

[1] Baptista M S. Cryptography with chaos. *Physics Letters A*, 1998, 240(1/2): 50–54.

[2] Wong W K, Lee L P, and Wong K W. A modified chaotic cryptographic method. *Comp Phys Comm*, 2001, 138(3): 234–236.

[3] Palacios A and Juarez H. Cryptography with cycling chaos. *Physics Letters A*, 2002, 303(5/6): 345–351.

[4] Wong K W. A fast chaotic cryptographic scheme with dynamic look up table. *Physics Letters A*, 2002, 298(4): 238–242.

[5] Wong K W, Ho S W, and Yung C K. A chaotic cryptography scheme for generating short ciphertext. *Physics Letters A*, 2003, 310(1): 67–73.

[6] Wong K W. A combined chaotic cryptographic and hashing scheme. *Physics Letters A*, 2003, 307(5/6): 292–298.

[7] Li S J, Mou X Q, Ji Z, and Zhang J, *et al*.. Performance analysis of Jakimoski–Kocarev attack on a class of chaotic cryptosystems. *Physics Letters A*, 2003, 307(1): 22–28.

[8] Li S J, Chen G R, Wong K W, and Mou X Q, *et al*.. Baptista -type chaotic cryptosystems: Problems and countermeasures. *Physics Letters A*, 2004, 332, (5/6): 368–375.

[9] Wong K W, Man K P, and Li S J, *et al*.. A more secure chaotic cryptographic scheme based on dynamic look-up table. *Circuits, Systems & Signal Processing*, 2005, 24(5): 571–584.

[10] Wei J, Liao X F, and Wong K W, *et al*.. Analysis and improvement for the performance of Baptista's cryptographic scheme. *Physics Letters A*, 2006, 354(1-2): 101–109.

[11] Jokimoski G and Kocarev L. Analysis of some recently proposed chaos-based encryption algorithm. *Physics Letters A*, 2001, 291(6): 381–384 .

[12] Álvarez G, Montoya F, Romera M, and Pastor G. Cryptanalysis of dynamic look-up table based chaotic cryptosystems. *Physics Letters A*, 2004, 326(3/4): 211–218.

[13] Chen Y and Liao X F. Cryptanalysis on a modified

Baptista-type cryptosystem with chaotic masking algorithm. *Physics Letters A*, 2005, 342(5/6): 389–396.

[14] Shanon C E. Communication theory of secrecy systems. *Bell Sys Tech J*, 1949, 28(4): 656–715.

[15] 李红达, 冯登国. 基于复合离散混沌动力系统的序列密码算法. 软件学报, 2003, 14(5): 991–998.
Li H D and Feng D G. Stream cipher algorithms based on composite nonliner discrete chaotic dynamical systems. *Journal of Software*, 2003, 14(5): 991–998.

[16] 胡汉平, 刘双红, 王祖喜等. 一种混沌密钥流产生方法. 计算机学报, 2004, 27(3): 408–412.
Hu H P, Liu S H, and Wang Z X, *et al*.. A Method for generating chaotic key stream. *Chinese Journal of Computer*, 2004, 27(3): 408–412.

[17] IEEE Computer Society. IEEE standard for binary floating-point arithmetic, ANSI/IEEE Std 754–1985 (August 1985).

葛　辛：女，1980 年生，硕士生，研究方向为混沌加密与保密通信.

刘粉林：男，1964 年生，教授，博士生导师，研究方向为图像处理与信息隐藏、信息安全.

芦　斌：男，1982 年生，硕士生，研究方向为信息隐藏与数字水印.

陈　娟：女，1982 年生，硕士生，研究方向为信息隐藏与数字水印.