😵 🎄 💐 Spring notes for CTF 🚩

Week 6

`Astrageldon`

2024-04-19

Contents

1 Misc

1.1 LSB Steganography Revelation

Disclaimer:　本节密集恐惧症患者慎入

本节主要讨论 LSB 隐写后的图片区别于隐写前图片的差异以及检测这种差异的（统计学）方法。准备一张正常的图片，这里选用以下图片（`src.jpg`）作为样例：

随机选取一个子块（这里取 128x128 大小），将子块中每一个像素点的 LSB 位伪随机地替换成 0 或者 1——这与隐写入秘密信息具有类似的效果。

`task1.py` (Github 🧐 🔗)

```python
import cv2, random, itertools
import numpy as np

def lsb_space_random(src):
    img = src.copy()
    img &= 0xfe
    return img | np.random.randint(0,2,img.shape)

src = cv2.imread("src.jpg")

h, w, c = src.shape

H, W = 128, 128

i, j = random.randint(0, (h-H+1)//H) * H, random.randint(0, (w-W+1)//W) * W
dst1 = src.copy()
dst1[i:i+H,j:j+W] = lsb_space_random(src[i:i+H,j:j+W])

print(i, i+H, j, j+W)
cv2.imwrite("dst1.png", dst1)
cv2.imwrite("src_lsb.png", (src[:,:,:]&1)*255)
cv2.imwrite("dst1_lsb.png", (dst1[:,:,:]&1)*255)
cv2.imwrite("src_block1_lsb.png", (src[i:i+H,j:j+H,:]&1)*255)
cv2.imwrite("dst1_block1_lsb.png", (dst1[i:i+H,j:j+H,:]&1)*255)
cv2.imwrite("src_random_lsb.png", (lsb_space_random(src[:,:,:])&1)*255)
```

Fig. 1 左为原始图像的 LSB 位，中为隐写了某个子块后图像的 LSB 位，右为整张图片隐写后的 LSB 位

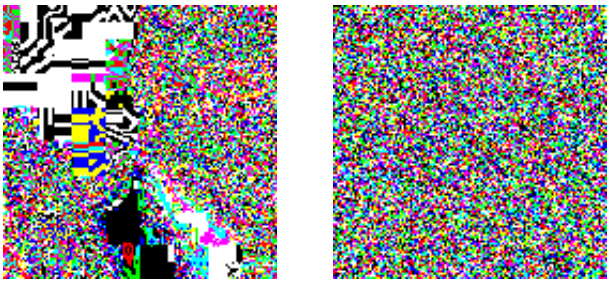


Fig. 2 左为目标子块的 LSB 位，右为对该子块隐写后的 LSB 位

比较 Fig. 1 左图与 `src.jpg`，可以发现一张图片的 LSB 位仍残留有原始图片的特征，而对比 1 左图与右图，这种特征在右图中意料之中地不复存在了，并且右图的像素点分布看上去更为混沌。因此，现在假设我们需要仅仅根据 `dst1.png` 找到这个被隐写过了的子块，我们就需要对图片中每一个子块进行某种检测，并要求这种检测——可以通过肉眼或者某种算法——能够区分出隐写前后的图片。

1.1.1 Plan A: Observation with Naked Eyes 👀

最为简单朴素的方法，就是将每个子块提取出来，用肉眼观测其混沌程度，然而，当子块比较小的时候，读者需要修炼出火眼金睛 👀 以接受五彩缤纷 🌈 的三味真火 🔥 的洗礼，为此笔者在 Fig. 3 中贴心地为读者准备了九九八十一道练就火眼金睛用的测试题（用“0”代表混沌，即随机的噪点；“1”代表秩序，也即从图片中提取出的 LSB 位）。

将这门本领练就至极致，方可看破这片混沌，让 LSB 精的伪装瞬间消弭于无形。

Fig. 3 九九八十一道火眼金睛练习题（16x16）

答案：0100010011011110101011000010000100100011111010001010011111000010010111101101010 11

1.1.2 Plan B: The χ^2 Test Attack! ⚔️

根据 🔗 维基百科，自由的百科全书 所述：

卡方检验（Chi-Squared Test 或 χ^2 Test）是一种统计量的分布在零假设成立时近似服从卡方分布（χ^2 分布）的假设检验。在没有其他的限定条件或说明时，卡方检验一般代指的是皮尔森卡方检定。在卡方检验的一般运用中，研究人员将观察量的值划分成若干互斥的分类，并且使用一套理论（或零假设）尝试去说明观察量的值落入不同分类的概率分布的模型。而卡方检验的目的就在于去衡量这个假设对观察结果所反映的程度。

而事实上，我们可以通过卡方检验来衡量一组数据观测得到的值 ξ_i 和理论期望的值 ξ_i^* 之间的差异，并在这个量度低于显著性差异值 α 拒绝零假设 H_0。

在正式介绍 χ^2 检验的概念之前，我们先给出几个引理。

Lemma 1.1.1 *若 X, Y 是两个相互独立的连续随机变量，密度函数分别为 $p_X(x)$，$p_Y(y)$，若 $Z = X + Y$ 的密度函数为 $p_Z(z)$，则*

$$p_Z(z) = (p_x * p_y)(z) = \int_{-\infty}^{\infty} p_X(z-t)p_Y(t)\mathrm{d}t$$

◁

为了得到上述公式，我们计算

$$\begin{aligned} F_Z(z) = P(X+Y \le z) &= \iint\limits_{x+y\le z} p_X(x)p_Y(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-y} p_X(x)\mathrm{d}x\right)p_Y(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{z} p_X(z-t)p_Y(t)\mathrm{d}t\mathrm{d}y \\ &= \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} p_X(z-t)p_Y(t)\mathrm{d}y\right)\mathrm{d}t \end{aligned}$$

两边对 z 求导即可得到欲证等式。

▷

Lemma 1.1.2 *若 $X \sim \mathrm{Ga}(\alpha_1, \lambda)$, $Y \sim \mathrm{Ga}(\alpha_2, \lambda)$，则 $Z = X + Y \sim \mathrm{Ga}(\alpha_1 + \alpha_2, \lambda)$。*

◁

显然

$$\begin{aligned} p_Z(z) = (p_X * p_Y)(z) &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\int_0^z (z-t)^{\alpha_1-1}e^{-\lambda(z-t)}\cdot t^{\alpha_2-1}e^{-\lambda t}\mathrm{d}t \\ &= \frac{\lambda^{\alpha_1+\alpha_2}e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\int_0^z (z-t)^{\alpha_1-1}t^{\alpha_2-1}\mathrm{d}t \\ &= \frac{\lambda^{\alpha_1+\alpha_2}e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}z^{\alpha_1+\alpha 2-1}\int_0^1 (1-t)^{\alpha_1-1}t^{\alpha_2-1}\mathrm{d}t \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}z^{\alpha_1+\alpha 2-1}e^{-\lambda z} \end{aligned}$$

其中最后一步是由于

$$\mathrm{B}(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

▷

Lemma 1.1.3 *若 $X \sim N(0,1)$，则 $Z = X^2 \sim \mathrm{Ga}\left(\frac{1}{2}, \frac{1}{2}\right)$。*

◁

$$\begin{aligned} P(Z \le z) = p(X^2 \le z) &= P(-\sqrt{z} \le X \le \sqrt{z}) \\ &= F_X(\sqrt{z}) - F_X(-\sqrt{z}) \end{aligned}$$

两边对 z 求导数，得到

$$p_Z(z) = \frac{1}{2} z^{-\frac{1}{2}} \left(p_X(\sqrt{z}) + p_X(-\sqrt{z}) \right)$$

代入 $p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ 即可。

▷

Theorem 1.1.1 设随机变量 $X_1,\ X_2,\ \cdots,\ X_n \sim N(0,1)$ 相互独立，则 $Y = \sum_{i=1}^{n} X_i^2$ 服从自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$，其密度函数为

$$p_n(y) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} & y > 0 \\ 0, & y \le 0 \end{cases}$$

也即 $Y \sim \chi^2(n) = \mathrm{Ga}\left(\frac{n}{2}, \frac{1}{2}\right)$。

◁

由前三个引理显见。

▷

Lemma 1.1.4 向 r 个盒子 $B_1, \cdots, B_r$ 中投掷 n 个球 $b_1, \cdots, b_n$，且 $P(b_j \in B_i) = p_i,\ i = 1, \cdots, r,\ j = 1, \cdots, n$ ($\sum_{i=1}^{r} p_i = 1$)，令 $\nu_i = \#\{b_j : b_j \in B_i\},\ i = 1, \cdots, r$，则有

$$n \to \infty, \qquad \sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j} \to \sum_{j=1}^{r} Z_j^2, \qquad where\ Z_i \sim N(0, \sqrt{1-p_i}), \quad E(Z_i^2) = 1 - p_i, \quad \mathrm{Cov}(Z_i Z_j) = -\sqrt{p_i p_j}$$

◁

由中心极限定理与简单的计算显见。

▷

Lemma 1.1.5 设 $g_1, \cdots, g_r \in N(0,1)$ *i.i.d.*，此外，

$$\mathbf{g} = (g_1, \cdots, g_r)^T, \qquad \mathbf{p} = (\sqrt{p_1}, \cdots, \sqrt{p_r})^T$$

则 $\mathbf{g} - (\mathbf{g} \cdot \mathbf{p})\mathbf{p}$ 与引理 *1.1.4* 中所述的 $(Z_1, \cdots, Z_r)$ 每个对应项都同分布。

◁

只需证

$$E\left(g_i - \sum_{l=1}^{r} g_l \sqrt{p_l}\sqrt{p_i} \right)\left(g_j - \sum_{l=1}^{r} g_l \sqrt{p_l}\sqrt{p_j} \right) = -\sqrt{p_i p_j}$$

$$E\left(g_i - \sum_{l=1}^{r} g_l \sqrt{p_l}\sqrt{p_i} \right)^2 = 1 - p_i$$

而这是显然的。

▷

至此，我们可以给出以下定理：

Theorem 1.1.2 ***(Pearson)***

$$\sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j} \to^d \chi^2(r-1)$$

◁

根据引理 1.1.5 我们可以做如下改写：

$$\sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j}} \right)^2 \to^d |\mathbf{g} - (\mathbf{g} \cdot \mathbf{p})\mathbf{p}|^2$$

而由于 $|\mathbf{p}|=1$，故 $\mathbf{g}-(\mathbf{g}\cdot\mathbf{p})\mathbf{p}$ 与 $\mathbf{p}$ 垂直。鉴于此，我们构造一个单位正交基 $\mathbf{p},\mathbf{p}_2,\cdots,\mathbf{p}_r$，$\mathbf{g}-(\mathbf{g}\cdot\mathbf{p})\mathbf{p}$ 在该基下的坐标为：

$$(0,g_2',\cdots,g_r')$$

不难证明 $g_2',\cdots,g_r'\in N(0,1)$ i.i.d.，因此 $(g_2')^2+\cdots+(g_r')^2\sim\chi^2(r-1)$，从而 $|\mathbf{g}-(\mathbf{g}\cdot\mathbf{p})\mathbf{p}|^2\sim\chi^2(r-1)$

▷

有了上述的定理 1.1.2，我们便有了进行 χ^2 检验的前置工具。现在，假设我们并不知道上述 p_i 而选择自己的一组 p_i°，有两个假设互斥地成立：

$$H_0:\quad \forall i,\ p_i=p_i^\circ$$
$$H_1:\quad \exists i,\ p_i\neq p_i^\circ$$

若零假设 H_0 成立，则根据定理 1.1.2，

$$\chi^2=\sum_{j=1}^{r}\frac{(\nu_j-np_j^\circ)^2}{np_j^\circ}\to^d\chi^2(r-1)$$

其中 ν_i 是我们实际测量到的数据。反之，若备择假设 H_1 成立，由于

$$\frac{\nu_i-np_i^\circ}{\sqrt{np_i^\circ}}=\frac{\nu_i-np_i+n(p_i-p_i^\circ)}{\sqrt{np_i^\circ}}=\underbrace{\sqrt{\frac{p_i}{p_i^\circ}}\frac{\nu_i-np_i}{\sqrt{np_i}}}_{\to^d N(0,(1-p_i)p_i/p_i^\circ)}+\underbrace{\sqrt{n}\frac{p_i-p_i^\circ}{\sqrt{p_i^\circ}}}_{\text{diverges to }\pm\infty}$$

故 $\chi^2\to+\infty$。

这也就是说，当零假设 H_0 成立时，n 趋向于无穷时，χ^2 会倾向于服从 $\chi^2(r-1)$ 分布；反之当备择假设 H_1 成立时，χ^2 趋向于正无穷。而通过对密度函数积分计算得到的 χ^2 值对应的累积概率 $P(X\leq\chi^2)$ 则可以作为拒绝 H_0 接受 H_1 或者拒绝 H_1 接受 H_0 的判据。

在有了 χ^2 检验这样一种强大的工具以后，我们回到原来的话题上来。用 $f_i,\ i=0,1,\cdots,255$ 来表示各种像素点出现的频数，不妨简单粗暴地假设各像素值在七个高位比特位上是均匀分布的，那么刻意进行 LSB 隐写以后，最低的一位比特位就会变得比原来均匀得多，基于此，我们预计 f_{2k} 的期望值是

$$f_{2k}^*=\frac{f_k+f_{k+1}}{2},\quad k=1,2,\cdots,128$$

计算 χ^2 的值：

$$\chi^2=\sum_{k=1}^{128}\frac{(f_{2k}-f_{2k}^*)^2}{f_{2k}^*}\sim\chi^2(r),\quad n=256,\ r=\frac{n}{2}-1=127$$

以及传说中的 p 值

$$p=1-\frac{1}{2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}\right)}\int_0^{\chi^2}e^{-\frac{x}{2}}x^{\frac{r}{2}-1}\mathrm{d}x$$

当 p 极其接近于 1 时，认为该子块被 LSB 隐写过，否则认为是正常的图片子块。值得指出的是，与理论预测相同，当子块大小不够大时（这意味着 n 值过小），χ^2 检验的效果会变差。

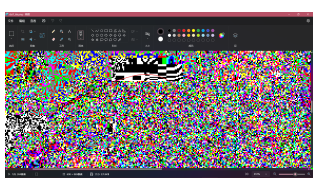
以下是对该统计学检验方法的一个 PoC（Proof of Concept）。

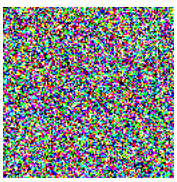
`task3.py` (Github 🧐 🔗)

```
import cv2, random, itertools, scipy
import numpy as np

def count(arr):
```

```
    ret = [0] * 256
    for i in arr:
        ret[int(i)] += 1
    return ret

def lsb_space_chi2(data0):
    data = data0.copy().flatten()
    bins = count(data)
    chi2 = 0
    for k in range(1,129):
        nexp = (bins[2*k-2]+bins[2*k-1])/2
        if nexp > 1:
            chi2 += (bins[2*k-1]-nexp)**2/nexp
    r = 127
    r2 = 0.5*r
    p = 1 - scipy.integrate.quad(lambda x: x**(r2-1)*np.exp(-0.5*x)/scipy.special
        .gamma(r2)/2**r2,0,chi2)[0]
    return chi2, p

dst1 = cv2.imread("dst1.png")

h, w, c = dst1.shape
H, W = 128, 128

candidates = []
candidates_capacity = 10

for i in range(0, h, H):
    for j in range(0, w, W):
        p = max(lsb_space_chi2(dst1[i:i+H,j:j+W,c])[1] for c in range(3))
        if len(candidates) >= candidates_capacity:
            candidates = candidates[:-1]
        candidates.append((i, j, p))
        candidates.sort(key = lambda x: -x[-1])

for candidate in candidates:
    print("\t%3d\t%3d\t%10.8f" % candidate)
```

```
384     128     0.99999995
  0     384     0.99997790
  0     512     0.99995744
128     384     0.99970638
128       0     0.99946186
256       0     0.98891486
384     384     0.98348694
128     128     0.95555590
256     384     0.89317160
768     512     0.00289893
```





结果是……符合预期的（嘻嘻）

1.1.3 Plan C: The Runs Test 🏃

根据 🔗 维基百科，自由的百科全书 所述：

The Wald – Wolfowitz runs test (or simply runs test), named after statisticians Abraham Wald and Jacob Wolfowitz is a non-parametric statistical test that checks a randomness hypothesis for a two-valued data sequence. More precisely, it can be used to test the hypothesis that the elements of the sequence are mutually independent.

由于维基百科简直是说的道理带着岛里的道莉和刀粒倒立着从道里买了几条稻鲤祷礼后回到家里被道理的妈叨厉了几声然后说了句“到哩”——有道理到家了，所以笔者不打算在此浪费笔墨 🖋️。PoC 如下：

task4.py (Github 🥹 🔗)

```python
import cv2, random, itertools, scipy
import numpy as np
from statsmodels.sandbox.stats.runs import runstest_1samp

dst1 = cv2.imread("dst1.png")

h, w, c = dst1.shape
H, W = 128, 128

candidates = []
candidates_capacity = 10

for i in range(0, h, H):
    for j in range(0, w, W):
        p = max(runstest_1samp(dst1[i:i+H,j:j+W,c].flatten()&1)[1] for c in range
            (3))
        if len(candidates) >= candidates_capacity:
            candidates = candidates[:-1]
        candidates.append((i, j, p))
        candidates.sort(key = lambda x: -x[-1])

for candidate in candidates:
    print("\t%3d\t%3d\t%10.8f" % candidate)
```

```
256     0    0.97788155
384   128    0.95118948
256   384    0.80900867
256   256    0.67576287
128     0    0.55501797
384   512    0.12972454
128   512    0.01684117
512   384    0.01673630
  0   512    0.00719960
768   512    0.00000000
```

效果不如 Plan B 捏 🥹 🥹 🥹

2 Maths

2.1 Several Number Theory Problems

以下问题节选自 Algebraic Number Theory, J.S. Milne 第二章的 Exercises。

2.1.1 $x^2+y^2=p \Leftrightarrow p\equiv 1 \pmod 4$

求证：$p\in\mathbb{P},\ p>2$，则下面三个条件等价：

a) $p\equiv 1 \pmod 4$

b) (p) 在 $\mathbb{Z}[i]$ 上可分

c) 关于 x,y 的方程 $x^2+y^2=p$ 在 $\mathbb{Z}$ 上有解

◁

(p) 在 $\mathbb{Z}[i]$ 上可分当且仅当 i 的极小多项式 i^2+1 在 $\mathbb{F}_p$ 上可分（$i^2+1=(i-i_1)(i-i_2)\Rightarrow(p)=(p,i-i_1)(p,i-i_2)$），此时 i 为 $\mathbb{F}_p$ 中阶为 4 的元素，于是 $4\mid p-1$。

现在设 (p) 在 $\mathbb{Z}[i]$ 上可分，$(p)=\mathfrak{p}_1\mathfrak{p}_2$，那么 $\mathfrak{p}_1=(x+yi),\ \mathfrak{p}_2=(x-yi)$，那么 $p=u(x+yi)(x-yi)=u(x^2+y^2)$，其中 $u\in\mathbb{Z}[i]$ 满足 $\exists v\in\mathbb{Z}[i],\ uv=1$，显然 $u=\pm1,\pm i$，因此必然有 $p=x^2+y^2$。反过来，若 $p=x^2+y^2$，那么 $(p)=(x+iy)(x-iy)$。

▷

2.1.2 $x^2+2y^2=p \Leftrightarrow p\equiv 1,3 \pmod 8$

求证：$p\in\mathbb{P},\ p>2$，则 $\exists x,y\in\mathbb{Z},\ x^2+2y^2=p\Leftrightarrow p\equiv 1,3 \pmod 8$。

◁

与上一个例子类似，只需证 α^2+2 在 $\mathbb{F}_p$ 上可分 $\Leftrightarrow p\equiv 1,3 \pmod 8$，左式成立当且仅当

$$\left(\frac{-2}{p}\right)=\left(\frac{-1}{p}\right)\left(\frac{2}{p}\right)=1$$

众所周知

$$\left(\frac{-1}{p}\right)\equiv(-1)^{\frac{p-1}{2}}\equiv 1 \pmod p \Leftrightarrow \frac{p-1}{2}\equiv 0 \pmod 2 \Leftrightarrow p\equiv 1 \pmod 4$$

$$\left(\frac{-1}{p}\right)\equiv(-1)^{\frac{p-1}{2}}\equiv -1 \pmod p \Leftrightarrow \frac{p-1}{2}\equiv 1 \pmod 2 \Leftrightarrow p\equiv -1 \pmod 4$$

此外，构造分圆域 $K=\mathbb{Q}(\zeta_8)$，考虑其 Frobenius 自同构 $\left(\frac{K/\mathbb{Q}}{p}\right)$，我们有

$$\left(\frac{K/\mathbb{Q}}{p}\right)(\sqrt{2})=\left(\frac{2}{p}\right)\sqrt{2},\qquad \left(\frac{K/\mathbb{Q}}{p}\right)(\zeta_8+\zeta_8^{-1})=\zeta_8^p+\zeta_8^{-p}$$

因此

$$\left(\frac{K/\mathbb{Q}}{p}\right)(\sqrt{2})=2\cos\left(\frac{p\pi}{4}\right)$$

即

$$\left(\frac{2}{p}\right)=\sqrt{2}\cos\left(\frac{p\pi}{4}\right),\qquad \left(\frac{2}{p}\right)=1\Leftrightarrow p\equiv\pm1 \pmod 8,\qquad \left(\frac{2}{p}\right)=-1\Leftrightarrow p\equiv 3,5 \pmod 8$$

取交集得

$$\left(\frac{-2}{p}\right)=1\Leftrightarrow p\equiv 1,3 \pmod 8$$

▷

2.1.3 $x^2+3y^2=p \Leftrightarrow p\equiv 1 \pmod 3$

求证：$p\in\mathbb{P},\ p>2$，则 $\exists x,y\in\mathbb{Z},\ x^2+3y^2=p \Leftrightarrow p\equiv 1 \pmod 3$。

◁

众所周知，当 $\mathbb{P}\ni a\equiv -1 \pmod 4$ 时

$$\left(\frac{-a}{p}\right)=\left(\frac{p}{a}\right)$$

上式在 $a=3$ 时当且仅当 $p\equiv 1 \pmod 3$ 时为 1。

▷

2.1.4 $x^2-dy^2=p$, when d is square-free?

作为对上述问题的一个拓展，当 y^2 前的系数 $-d$ 为不含平方因子的整数时，我们初步地知道求解 p 的范围的问题照例可以转化为求使得

$$\left(\frac{d}{p}\right)=1$$

的 p 值的问题，至于具体的求解方式且听下回分解（挖坑）

什么？求 x,y？那是佩尔方程才需要管的东西 🧐（出门右转：CSDN🔗）