統計與多變量分析應用於旁通道攻擊

陳君朋

摘要

旁通道攻擊論文在 1996 年首先被提出後，過去二十年該領域便受到各界矚目。過去沒人意識到，即使演算法在數學上無法找尋到弱點，卻因加密裝置運作洩漏物理訊息，導致裝置依然存在資安風險。分析人員可利用利用加密過程的功率消耗、電磁輻射，以及其他手法推算出正確金鑰。因為這些安全顧慮，有許多研究學者提出對應的防禦措施；而這些方式普遍的指導原則，就是如何防止關鍵敏感資訊被辨識。本篇所關注的是近十年來，相關統計學工具在旁通道分析中常見應用，及如何使用統計工具，快速分析資訊洩漏所在位置與程度。

I. 前言

Whitfield Diffie與 Martin E. Hellman兩位學者，在 1976 年提出 New Directions in Cryptography [1] 之後，公開金鑰密碼學 (Public-Key Cryptography, PKC) [2] 在過去四十年被全世界廣泛使用，最具代表性的是 RSA 與橢圓曲線密碼學 (Elliptic Curve Cryptography, ECC)。而這兩個演算法的安全性，是基於質因數分解與橢圓曲線離散對數問題。

Peter W. Shor 在 1994 發表 Algorithms for Quantum Computation: Discrete Logarithms and Factoring [3] (現稱為 Shor's algorithm) 表示，當有足夠好的量子電腦，解離散對數問題與分解質因數，只需要多項式時間即可完成。而現今量子電腦的研究成果被全世界關注，預計未來 10-15 年發展到一定程度後，將對現行公鑰密碼系統有莫大的威脅 [4] 。也就是說：量子電腦發展越蓬勃，現行的公鑰密碼系統就越不安全。

有鑑於此，美國國家標準暨技術研究院 (National Institute of Standards and Technology, NIST) 2017 年底開始，對全世界徵選抵抗量子電腦攻擊的演算法，現統稱為後量子密碼學 (Post-Quantum Cryptography, PQC) [5]。今年(2019)徵選進入競賽的第二回合，剩下約 26 個候選演算法 [6]，可預期進入第三回合的演算法剩下十個左右，並預計在 2022-2024 年左右公布徵選的標準[7]，以替換目前被廣為使用的 RSA 與 ECC。

在相同安全條件下，ECC 所需要的處理器資源遠小於 RSA，近年來物聯網越來越蓬勃發展，有限計算量的硬體裝置，彼此溝通的安全性也越受重視。因此利用 ECC 增進無線感測網路 (Wireless Sensor Network, WSN) [8]、射頻辨識 (Radio-Frequency Identification, RFID) [9]，物聯網 (Internet of Thing, IoT) [10] 的安全應用與研究，過去十年至今仍持續廣泛討論中。

任何密碼系統的硬體實作，都會面臨旁通道攻擊 (Side-Channel Attack, SCA) 等相關問題 [11]。該研究領域自 Paul Kocher 於 1996 年首先提出，並成功藉由演算法於硬體運行時，洩漏的功率消耗曲線分析出密鑰 [12]-[13]。無論是處理器的計算時間、功率消耗、電磁輻射 [14]，或裝置運行時產生的聲音[15]，都可能洩漏密鑰的相關資訊。甚至整合射頻與數位邏輯的系統晶片(System on a Chip, SoC)，藉由接收射頻相關無線通訊訊號，反推數位邏輯加密運算時洩漏的雜訊，進而取得密鑰都成為可能 [16]；簡單的說，硬體裝置單純只有加解密的演算法，實作時卻沒考慮抵抗旁通道分析，安全性仍然不足。

近兩年左右，抵抗旁通道分析的後量子演算法硬體實作，正逐漸發表在頂尖研討會與期刊 [17]；應用機器學習等統計相關工具與演算法，以增進旁通道分析能力也逐漸被重視 [18]-[19]；NIST 後量子標準演算法的候選者，在有限資源的硬體裝置實作比較，也在今年被提出 [20]。可預見未來五到十年內，相關研究與應用將被廣泛討論，後量子演算法在硬體實作的研究也越來越重要。

在加解密演算法設計上，需要許多代數 (Algebra) 相關知識得以了解演算法原理；而在旁通道分析實作上，許多地方仰賴於統計學 (Statistic) 或多變量統計分析 (Multivariate Statistical Analysis) 相關工具。因此在本篇文章中，在第一部份將討論關於基礎統計的相關知識，而第二部分將著重在統計運用在旁通道分析的相關技巧。

II. 基礎統計學 [21-24]

（一） 連續隨機變數

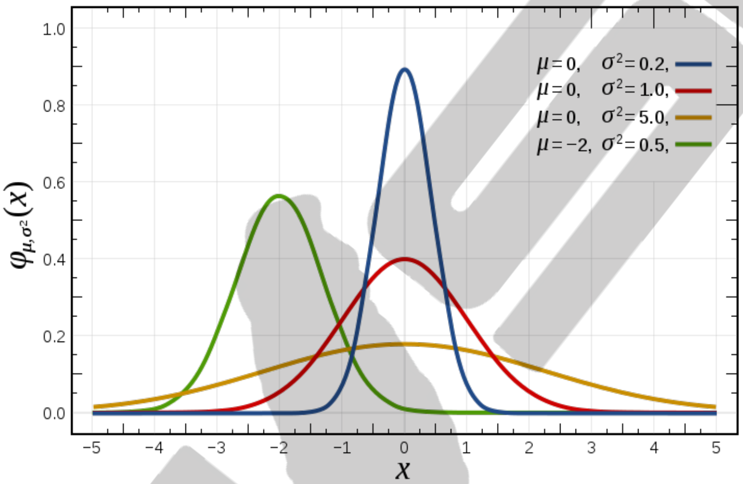
隨機變數(random variable)是一個函數，其意義在於：由狀態空間映射到一實數，通常以 $X(w) = x$ 表示。其中 w 表示狀態空間的可能出象，x 表示隨機變數的實現值(有時亦稱作隨機變量)。而連續隨機變數 (continuous random variable) 表示這函數可能實現值範圍於任何實數，並且個數為無限且不可數。而隨機變數 X 的期望值 (Expectation, Expected Value)，可定義

為：$\mu = \mathrm{E}(X)$，而隨機變數 X 的變異數(variance) 定義為：$\sigma^2 = Var(X)$，並 $Var(X) = E(X^2) - [E(X)]^2$，而標準差 (Standard deviation)為 σ 。

（二） 常態分佈

常態分佈 (Normal distribution) 又稱作 Gaussian distribution。當 X 為期望值 $E(X) =\mu$, 變異數 $Var(X) =\sigma^2$的常態隨機變數，其機率密度可寫成：

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



圖一、不同參數對所對應到的不同常態分佈曲線

圖一展示了常態分佈的鐘形曲線，圖的橫軸表示隨機變數的數值，而縱軸代表相對應的出現機率。不同顏色表示不同的機率密度函數 (Probability Density Function, PDF)。而期望值範圍為 $-\infty < \mu < \infty$，而標準差 $\sigma > 0$，常態分佈一般記為 $X \sim N(\mu, \sigma^2)$

由圖可見，常態分佈並沒有上下界，整個實數系上任一點出現的機率均不為零。事實上，特定的機率分佈是一個群體，並需要一些參數來表示群體中的不同個體，例如常態分佈中的 μ 和 σ。如圖一所示，微調參數可以得到樣貌相似的不同曲線，而曲線下面積為 1:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\,dx = 1$$

除了 PDF，另一個重要的函數為累積分佈函數(Cumulative Distribution Function, CDF)。PDF 與 CDF 均完整描述了一個機率分佈的行為，並且兩者

之間為一對一對應，只要給定兩者之一，我們就能藉由微積分得到另一個。PDF 曲線下面積為 1，而 CDF 會在隨機變量數值趨近無窮大時收斂到 1。

圖一紅線部分的 $\mu = 0$ 和 $\sigma = 1$ 稱之為標準常態隨機變數 (Standard normal random variables) Z，也就是將常態隨機變數 X 標準化。如：

$$Z = \frac{X - E(X)}{\sqrt{Var(X)}} = \frac{X - \mu}{\sigma}$$

只要將上式代入常態分佈的 PDF，即可發現期望值與標準差分別變成 0 與 1。標準常態分佈的用途廣泛，舉例而言，不直接靠積分計算常態分佈的 CDF 數值，而是先轉成標準常態分佈之後再查 CDF 數值表。 這是因為常態分佈的定積分難以在缺乏計算機的輔助下計算，甚至為了節省資源，連計算機也仰賴查表。

（三） 其他分佈

若有一序列的隨機變數$X_1,\ X_2, \dots,\ X_n$相互獨立，並且來自相同分佈 (independent and identically distributed, i.i.d.)，則稱之 i.i.d. 隨機變數。表示這個數為 n 的隨機變數彼此相互獨立，並且每個變量的機率分佈都相同。令 $X_1,\ X_2, \dots,\ X_n \sim N(\mu, \sigma^2)$ 為獨立相同分佈 (i.i.d.) 的n個常態隨機變數。則

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

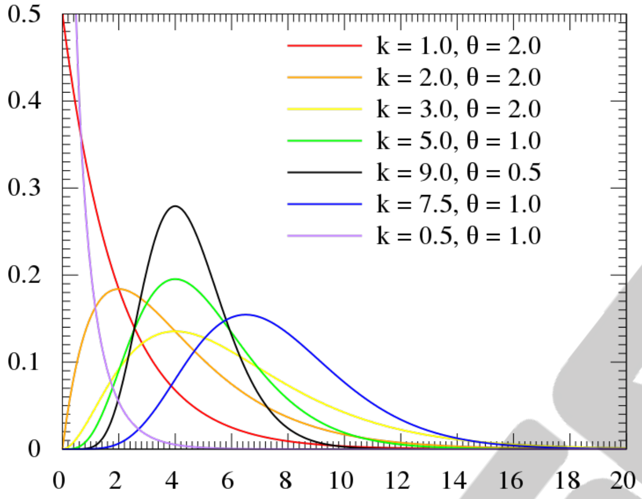
遵循常態分佈，期望值與標準差分別為 $\mu_{\bar{X}} = \mu,\ {\sigma_{\bar{X}}}^2 = \sigma^2/\text{n}$ 此外，可以將多個標準常態隨機變數，依照特定的組合，建構出新的機率分佈。舉例而言，卡方分佈就是多個 i.i.d. 的標準常態隨機變數的平方和。

同理，當 $Z_i = \frac{X_i - \mu}{\sigma}$ 亦為 i.i.d.且

$$\sum_{i=1}^{n} Z_i = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right);\sum_{i=1}^{n}\chi^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

卡方分佈一般表為 $\chi^2(n)$，其自由度為 n，該分佈是用來估計變異數的重要統計模型。此外，結合卡方分佈及常態分佈，我們可以建構出另外兩種機率分佈，分別為t分佈與 F 分佈。

當常態分佈的母體中所抽樣的樣本數夠多，那麼這些樣本的表現就足夠代表母體的性質。但如果樣本數不多，可採用 t 分佈或稱學生 t-分布（Student's t-



圖二、Gamma 分佈曲線

distribution）呈現結果。換句話說，當小樣本不足以代表母體性質，或如果母體分佈不為常態，則抽樣出來的分配也不是常態，而是依母體決定。實際的情況將隨『自由度 (degree of freedom) ν』變動，自由度越大則越近似常態分配。令 Z 為一標準常態隨機變數，而 W 係一自由度為 ν 的 χ^2 隨機變數。若 Z、W 獨立，則具有 t-分佈的隨機變數 T 可定義為：

$$T = \frac{Z}{\sqrt{\frac{W}{\nu}}}$$

並且其自由度為 ν，以 $t(\nu)$ 表示之。

令 W_1, W_2 為兩相互獨立之 χ^2 隨機變數，其自由度分別為 ν_1, ν_2，則則具有 F-分佈的隨機變數 F 可定義為：

$$F = \frac{\sqrt{\frac{W_1}{\nu_1}}}{\sqrt{\frac{W_2}{\nu_2}}}$$

其分子自由度為 ν_1，分母自由度為 ν_2，以 $F(\nu_1, \nu_2)$ 表示。

T 分佈多用於 Welch’s T-test，而卡方分佈其實 Gamma 分佈中的一個特殊集合。為此可簡單列出 Gamma 分佈的定義。對於一隨機變數 X，其遵循 Gamma 分佈的充要條件為 PDF 具備以下形式:

$$f(x) = \frac{\int_0^x x^\alpha e^{-\frac{x}{\beta}}\, dx}{\beta^\alpha \Gamma(\alpha)}, \qquad \Gamma(\alpha) = \int_0^\infty y^\alpha e^{-y}\, dy$$

，其中 $\alpha > 0$ 為形狀參數，而 $\beta > 0$ 則為尺度參數，一般以 $\Gamma(\alpha, \beta)$ 或

Gamma $(\alpha,\ \beta)$ 表示之。$\chi^2(\nu)$ 實為 Gamma 分佈中 $\alpha = \nu/2$ 且 $\beta = 2$ 的族群。圖二展示了不同參數的 Gamma 分佈的不同面貌。Gamma 隨機變量的取值下界為零。

（四） 中央極限定理 (Central Limit Theorem, CLT)

令 $X_1,\ X_2, \ldots,\ X_n$ 為 i.i.d. 的隨機變量，$E[X_i] = \mu$ 且 Var $(X_i) = \sigma^2 < \infty$。
定義以下函數:

$$U_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

則當 n 趨近於無窮大時，U_n 的 PDF 收斂為標準常態分佈的形式，意即對於任意 u 而言

$$\lim_{n\to\infty} P(U_n \le u) \approx \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}\, dt$$

U_n 的極限分配為標準常態 $N\,(0, 1)$

中央極限定理是機率中最重要的定理之一，由於它適用於任何的機率分佈，並表示當抽樣的樣本數夠大，則抽樣分佈 (sampling distribution) 會非常接近常態分佈，因此又稱漸進常態分佈。

至於「樣本數夠大」該如何界定，顯然對於不同種的機率分佈與不同情況，能夠套用所需的樣本數不盡然相同。在瞭解上述機率分佈及中央極限定理後，將介紹如何做出可靠的估計。

（五） 估計 (Estimate)

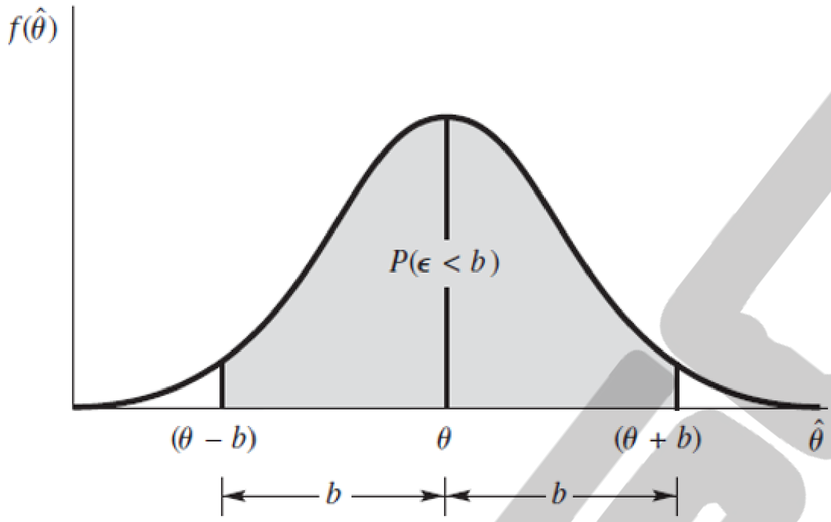
假設我們對某特定大群體的某參數 θ 感興趣 (例如: 北市生育率、全台 35 歲以下青年的總統大選投票率…等)。一般而言，由於該群體相當龐大，不可能以普查方式調查群體的每一份子，因此做法就是針對該大群體進行抽樣。

透過統計相對小的樣本而得到一個估值 $\hat{\theta}$，用以估計實際參數 θ。這類型的單一統計數值 $\hat{\theta}$ 稱作點估計。必須提醒的是，因為抽樣本身也是隨機的，在不同實驗中我們抽樣的對象、甚至樣本數都會不一樣，因此統計所得道的點估計 (Point Estimation) $\hat{\theta}$ 也會不同。因此，$\hat{\theta}$ 在本質上就是個隨機變數。

理想上，點估計 $\hat{\theta}$ 的期望值，要恰好是我們想估計的對象 θ 才行:

$$E\left[\hat{\theta}\right] = \theta.$$

滿足此條件的點估計 $\hat{\theta}$，我們稱之為不偏估計(unbiased estimator)，反之則稱為



圖三、於常態分佈曲線中的區間半徑 b

偏誤估計(biased estimator)。

在之前提到只要樣本數夠大，中央極限定理是個強而有力的工具，它使我們能夠用性質優異的常態分佈，來描述抽樣得來的參數估計，即使不確定原來單獨樣本的機率分佈為何。

若原機率分佈為丘型曲線 (mound-shaped curve)，只要樣本數 n 大於 30，基本上即適用中央極限定理。另外，若原機率分佈為二項分佈，則伯努力試驗參數估計 $\hat{p}$ 可近似為常態分佈的條件為 n 大於 p 與 $1-p$間比例的 9 倍，其中 p 與 $1-p$ 兩者大的當分子，較小的置於分母。當然，樣本數越大，常態分佈就能描述得更貼近真實。

另外，區間估計與點估計的不同之處在於，點估計方法將統計得到的估值 $\hat{\theta}$ 來估計未知參數 θ；而區間估計則是進一步延伸，根據 $\hat{\theta}$ 得到一個區間。當我們以此區間來估計 θ，代表我們認為 θ 落在區間內。由於 $\hat{\theta}$ 本身的隨機性，θ 仍有 α 的機率落在區間以外，故 α 稱為錯誤率。舉例來說，閉區間估計常以 $\hat{\theta}$ 為中心、半徑為 r 展開，此時 $\alpha = \Pr\{|\hat{\theta} - \theta| > \mathrm{r}\}$，而 $(1-\alpha)$ 即所謂的信賴係數 (confidence coefficient) 或者信心水準 (confidence level)。若我們重複實驗 k 次共得到 k 個區間，則約有 $\mathrm{k}(1-\alpha)$ 個區間可以涵蓋到真實的參數 θ。顯然區間估計比點估計更具說服力，因為區間估計已將點估計的隨機性納入考量。

錯誤率 α、區間半徑 b，以及樣本數 n，三者互為取捨關係。我們希望能壓低錯誤率，即使是運氣較差的抽樣樣本也能得到有效的區間估計；也希望壓

低區間半徑，因為那意味著估計較準確，同時給我們帶來較多資訊 (當然可以選擇無限大的區間半徑，然後 θ 必定落在區間內，也就是 $\theta \in \boldsymbol{R}$，然而這樣的區間卻是毫無價值。

參考圖三可定義點估計偏差 $\varepsilon = |\hat{\theta} - \theta|$，假使我們今天的抽樣得到較大的估計偏差 $\varepsilon = b$，如 $\hat{\theta} = \theta - b$，則選用小於 b 的區間半徑，將使得 θ 落在區間外，提高錯誤率。當個別樣本遵守常態分佈，樣本數越大時採樣分佈的標準差就越小，當然此敘述適用於任何分佈。當點估計的標準差下降，可以預期估計偏差就越小，表示相同區間半徑下，可以得到較低的錯誤率。樣本數足夠多的時候，的確可解決無法同時壓低錯誤率與區間半徑的問題，然而此舉也會使得抽樣程序變得更複雜，無論是時間或金錢成本都會增加。

因此，常見做法是預先設定目標的信心水準 (1-錯誤率)，再依此在精確度與抽樣成本間做權衡。在下一節的假設檢定流程中，我們也使用類似的策略。

（六） 假設檢定 (Hypothesis Testing)

假設工廠機台 A 生產的 100 個產品當中有 15 個瑕疵品，而根據之前制定的規則，可容忍的不良率為 10%。若希望決策的錯誤率在 0.1 以下，請問是否該將機台 A 送修? (註: 此不良率為針對個別產品而言，即伯努力試驗參數 p)

這是個經過簡化但不失貼切的實例。令 p 為機台 A 的不良率。雖然生產了 15% 的瑕疵品，卻無法斷定不良率是否真的超過 10%。就像擲銅板一樣，就算擲 100 次中只有 42 次正面，也無法因此就斷言它是個不公平的銅板。於是需要其他方法來協助我們做決策，而這方法就稱為假設檢定。

假設檢定的流程有以下四項組成要素:

I. 虛無假設 (Null Hypothesis, H_0)
II. 對立假設 (Alternative Hypothesis, H_1)
III. 檢定統計模型 (Test Statistics)
IV. 拒絕區間 (Rejection Region, RR)

一般而言，對於虛無假設，將其設定為本次檢定希望否決的對象；換句話說，希望對立假設成立。至於原因為何，須從假設檢定的精神去理解，以下沿用機台維修的例子做說明。

根據擬定的標準，若一台機器產出不合格產品的機率超過 10%，表示不良率偏高，必須進入維護週期。因此，機台 A 的不良率是否超過 10%，就是我們下

決策所需的資訊。由此，我們便可得到本次假設檢定的兩個要素:
令 p 為機台 A 不良率，則虛無假設 $H_0: p = 0.1$，對立假設 $H_1: p > 0.1$。

雖然無從得知機台 A 的實際 p 值，依然可計算出 p 為 0.1 時，今日觀察結果 (15%瑕疵) 的發生機率。如果發現假設 $p = 0.1$ 會導致已發生的事件幾乎不可能發生，我們便傾向相信假設錯誤。

程序就是：提出假設⟶得到不合理的情況⟶否決該假設。

根據中央極限定理，當樣本數夠大，取樣分佈會趨近於常態分佈。由於機台生產每一產品均可模擬為一次伯努力試驗，故單日產出的良莠可視為二項分佈。於之前提到，足以套用中央極限定理的二項分佈樣本數門檻為:

$$\text{樣本數 n} \times \max(\frac{p}{1-p}, \frac{1-p}{p})$$

本題中，此門檻為 9 * (0.9/0.1) = 81，小於樣本數 100，故中央極限定理於此處適用，機台 A 生產的瑕疵品數量之機率分佈可用鐘形曲線來近似，這便是檢定統計模型的概念。對照標準常態分佈 CDF 表可得:

$$\phi\left(\frac{0.15-0.1}{\sqrt{\frac{0.1\cdot 0.9}{100}}}\right) = \phi(1.667) \approx 0.9522$$

至於拒絕區間 RR，概念與前面所介紹的區間估計有些相似。假定我們預先決定拒絕發生機率最低的 5%事件，換言之，只要發生機率最低的 5%的事件都發生了，我們便相信一開始設定的虛無假設錯誤，因此拒絕虛無設。在本題中的確拒絕了虛無假設，原因是觀測值 0.15 落在常態分佈曲線最右側的 5%以內，此區域即為拒絕區間。

到此可能有所質疑: 常態分佈最右側 5%並非發生機率最小的 5%。此處的 95% 僅使用只有上界的單邊區間，而非大家平時熟悉有上下界的閉區間。事實上，使用何種區間，一般決定於對立假設的屬性。

依照慣例，我們的虛無假設通常為 $\theta = \theta_0$ 的形式，其中 θ 為數值未知的統計量，而 θ_0 則是我們對於該統計量的估值或假設值。我們額外以 θ'表示抽樣結果。此時，對立假設可有三種型式:

1. $\theta > \theta_0$
2. $\theta < \theta_0$
3. $\theta \neq \theta_0$

舉第一種形式為例，我們希望得到$\theta > \theta_0$的結果，因此不在意 θ_0 左側的情況。就意義上來說，θ' 小於 θ_0 就和 θ' 只略大於 θ_0 或恰約等於 θ_0 一樣，當進行假設檢定都會得到相同的結果: 無法拒絕虛無假設。因此拒絕區間為右側 5%的面積(單翼)。同理，若對立假設為上述的第二種形式，便改用最左側 5%的面積作為拒絕區間(單翼)。只有當對立假設為第三種型式時，95%區間才使用閉區間，此時拒絕區間就是兩側各 2.5%的面積(雙翼)。以數學形式表達統計模型如下:

$$Z = \frac{\theta' - \theta_0}{\sigma_{\theta'}}$$

拒絕區間:

$\theta > \theta_0$: $\mathrm{RR} = \{Z > Z_\alpha\}$ $\quad (upper - tail\ RR)$

$\theta < \theta_0$: $\mathrm{RR} = \{Z < -Z_\alpha\}$ $\ (lower - tail\ RR)$

$\theta > \theta_0$: $\mathrm{RR} = \{|Z| > Z_{\alpha/2}\}$ $\ (two - tailed\ RR)$

其中 $\mathrm{P}(Z > Z_\alpha) = 1 - \alpha$，而 α 便是拒絕域面積。

(註: 亦有些文獻使用的數學符號為 $\mathrm{P}(Z < Z_\alpha) = \alpha$，與本文不同。)

根據前面的敘述可了解，為何通常希望拒絕虛無假設? 以$\theta > \theta_0$ 的對立假設為例，「無法拒絕虛無假設」及意味著「無法斷定 θ 是否大於θ_0」，也就是說，本次檢定以無結論收場。當然，我們可以下修 θ_0，使得觀察到的 θ'與θ_0離遠一點，就有機會拒絕虛無假設。

然而，回到機器維修的例子，θ_0: $p = 0.1$顯然是個程序標準，當然不能根據不同日子、不同機器的情況任意替換。倘若今日 100 個產品中的瑕疵品只有 14 個而非 15 個，即 $\theta' = 0.14$，並且我們決定以最右側 5%作為拒絕區間，那麼便無法拒絕虛無假設。以另一種方式來說，「在 5%拒絕區間下，就算觀察到 $\theta' = 0.14$，我們似乎也無法因此斷定 $\theta > 0.1$。也許只是機台 A 今天運氣差一點罷了。」。當然有另外一種情況是，無法確定機台 A 的不良率是否已高達 0.18，或許可能今天運氣好一些，僥倖沒通過檢定。(通過檢定相當於必須進入維護週期。)

雖然區間估計與假設檢定分開介紹，兩者本質上幾無二致，差別只在描述對象的不同。為了方便比較與說明，假定我們持有夠多的樣本數，滿足中央極限定理的適用條件，均以 95%作為決策依據，並一律使用閉區間。符號使用上，我們依舊以 θ 為數值未知的統計量，θ_0 是對於該統計量的估值或假設值，θ'表示抽樣結果。

區間估計的概念如下:

不曉得 θ 身在何處，但根據數學理論，我們觀測到的 θ' 是個隨機變數，並且以常態分佈的形式以 θ 為中心展開來。理論上，只要我們運氣別太差(或

太好)，θ' 都不會離 θ 太遙遠，根據常態分佈的性質，有 95%的情況抽樣得到的 θ' 與 θ 差距，在兩個標準差之內。這就意味著當我們以兩倍標準差作為區間估計的半徑，這區間有 95%的機率包到實際的 θ。因此可說作為
θ 的估計，區間 $[\theta' - 2\sigma, \theta' + 2\sigma]$ 具備 95%的信心水準。
(請注意: 由於 θ 為定值，此處「95%的機率」源自於抽樣的不確定性)

假設檢定的概念則是:

不曉得 θ 身在何處，但根據數學理論，我們觀測到的 θ' 是個隨機變數，並且以常態分佈的形式以 θ 為中心展開來。理論上，我們運氣太差(或太好)的機率相當小，θ'不該離 θ 太遙遠。因此若猜了 $\theta = \theta_0$ 後卻發現， θ'超出以 θ_0 為中心的常態分佈，有兩倍標準差之外(太遙遠)，因為發生的機率並不高，因此表示 θ_0 為不合理的猜測，合理拒絕 $\theta = \theta_0$ 的虛無假設；反之，如果 θ_0 猜測得宜，該 θ' 無法支持拒絕虛無假設。(事實上可驗證，所有合理猜測的集合就是 $[\theta' - 2\sigma, \theta' + 2\sigma]$。)

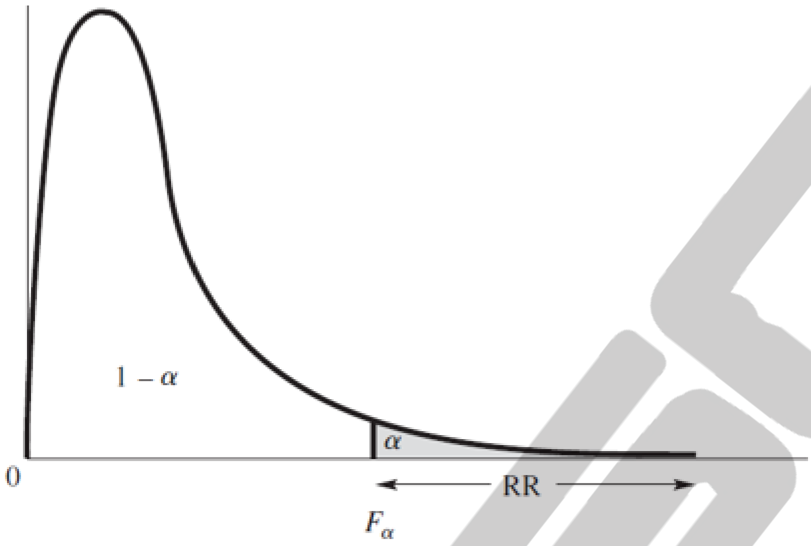
區間估計說明 θ 應該在 θ' 的區間裡，假設檢定則進一步說明 θ 應該在 θ' 的區間裡，倘若有個 θ_0 不在區間裡，我們合理斷定他並非 θ。

比較後要討論的是拒絕區間的選擇。如何決定拒絕區間的大小? 記得修機器的例子中曾提到，當認為一天產出 14%的瑕疵品，不足以肯定不良率超過 10%，事實上有可能實際上不良率已經達到 18%之高；另一方面，在認為 15%瑕疵品的現況，正催促儘快維修機器的同時，殊不知機台 A 今天真的只是運氣不佳而已。

對於以上兩種情形，我們分別稱之為型 2 錯誤與型 1 錯誤。型 1 錯誤 (Type I error)，指的是虛無假設為真，卻誤拒絕之的情況；型 2 錯誤(Type II error)，則是虛無假設為偽(對立假設為真)，卻沒拒絕虛無假設的情況。(請注意，虛無假設 $\theta = \theta_0$ 為真與否，需視情況而定，此處等號並不強硬限制左右數值必須確切相等。) 顯然，選擇拒絕區間界線的設立點，即是在型 1 錯誤與型 2 錯誤之間作取捨。此消彼長。因此如同之前所述，概念上，我們可以在做假設檢定之前，根據決策的保守程度，預先設定型 1 錯誤的最高容忍範疇(拒絕域大小)。

另外，有一種方式是引進更多資訊來給不同的錯誤型態乘上不同權重。比方說，修機器的機會成本大過不良率略高於 10%的機台帶來的負面影響，那麼，型 1 錯誤一旦發生，會給工廠造成的更多額外的負擔，相較之下工廠對型 2 錯誤的容忍度較高。我們的決策方式便由「降低誤判機率」轉移至「降低成本」，因此選擇將拒絕域的下界右移，以壓低型 1 錯誤的出現機會。又或者可考慮連續抽樣三天，以期分散極端抽樣的影響，這部分在旁通道分析領域就有類

似做法。



圖四、F 分佈之 PDF 與 F 檢定拒絕區間(RR)

以上介紹了中央極限定理適用的狀況，但是當中央極限定理不再適用，唯一的差別是必須以其他的檢定統計模型取代常態分佈，其餘概念均與原先相同。倘若今天要檢定的對象是變異數 σ^2，了解其是否為某個常數 ${\sigma_0}^2$，我們選用 χ^2分佈(卡方分佈)作為檢定統計模型:

$$\chi^2 = \frac{(n-1)S^2}{{\sigma_0}^2}$$

拒絕區間:

$\sigma^2 > {\sigma_0}^2$: $\mathrm{RR} = \{\chi^2 > \chi^2_{\alpha}\}$ $\quad (upper - tail\ RR)$

$\sigma^2 > {\sigma_0}^2$: $\mathrm{RR} = \{\chi^2 < \chi^2_{1-\alpha}\}$ $\quad (lower - tail\ RR)$

$\sigma^2 > {\sigma_0}^2$: $\mathrm{RR} = \{\chi^2 > \chi^2_{\alpha/2}\} \cup \{\chi^2 < \chi^2_{1-\alpha/2}\}$ $\quad (two - tailed\ RR)$

若對象一樣是期望值 μ 、伯努力試驗參數 p、$\mu_1 - \mu_2$或者 $p_1 - p_2$，但樣本數不夠大時，則檢定統計模型必須使用 t 分佈(以 μ 為例) :

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}}$$

比較特殊的是當檢定對象為兩族群變異數 ${\sigma_1}^2$ 與 ${\sigma_2}^2$是否相等時，檢定統計模型採用 F 分佈:

$$F = \frac{{S_1}^2}{{S_2}^2}$$

以上三種檢定僅列出模型數學式，涵義與推導過程不多做論述。最後要介紹的數值稱為 P-value (p 值) (勿與伯努力參數 p 混為一談)。P-value 的定義為「下一

次抽樣結果，比本次更有利於拒絕虛無假設的機率」。15% 瑕疵品落在鐘形曲線上的位置為 $\phi(1.667) \approx 0.9522$，故抽樣的 p-value 為$1 - 0.9522 \approx 0.448$，換言之，若 P-value 小於拒絕域的面積，表示本次抽樣結果支持拒絕虛無假設。

III. SCA 常見的統計學指標 [24-26]

（一） NICV (Normalized Inter-Class Variance for Detection of Side-Channel Leakage)

過往檢定硬體裝置的洩漏資訊，用以分析對旁通道攻擊是否具備足夠抵抗能力，其程序往往曠日廢時。需要將待測裝交給專門實驗室做檢測，經過一連串檢驗設計與分析，才能得到最終結果。

為加速檢驗流程，有許多檢測手法因應而生，而 NICV 便是其中一種檢測數值指標。進行高強度的差分能量攻擊 (Differential Power Analysis, DPA) 需要大量的能量消耗變化曲線 (Power Trace)；而一條能量曲線，又仰賴數以百計的資料點繪製而成。如此大量資料的影響，除了儲存空間的浪費，檢測時間也必定非常耗時。

可想而知，當中多數的資料點參考價值並不大，能量消耗的資訊洩漏多寡係因資料點而異，我們只在意洩漏資訊最多的地方。即使百萬個資料點當中，只要有一個點洩漏了金鑰資訊，攻擊者就有機會破解該裝置。NICV 指標的誕生，即蘊含著揪出所有潛在機密資訊洩漏點的精神。去除非潛在高風險的資料點，也是一種資料壓縮的方式。

NICV 有以下幾項特點:

- 僅需能量曲線與明文密文對即可分析，並不需對該加密裝置有深入了解。
- 不須建立能量模板。
- 只能得到潛在高風險的位置，以降低攻擊或檢測所需付出的心力與時間。如需取得金鑰資訊仍需實際進行攻擊。
- 可根據該指標給出的結果加以推測更多裝置細部資訊，例如資訊洩漏模型 (leakage model)。

在接續的討論中，我們將陸續提到這些性質。而下表列出本節所使用的符號:

X	(部分)明(密)文 (多指 1 位元組)
Y	能量曲線(可視為能量資料點的有序集合)上的一個資料點(可為實錄或仿真結果)
K	(部分)猜測金鑰 (多指 1 位元組)

K^*	(部分)實際金鑰 (多指 1 位元組)
L	資訊洩漏模型，輸入為部分明(密)文與金鑰，或者只有部分明(密)文
ρ	相關係數
c	常數

首先，NICV 的定義如下:

$$NICV = \rho^2[E[Y|X];Y] = \frac{Var[E[Y|X]]}{Var[Y]}$$

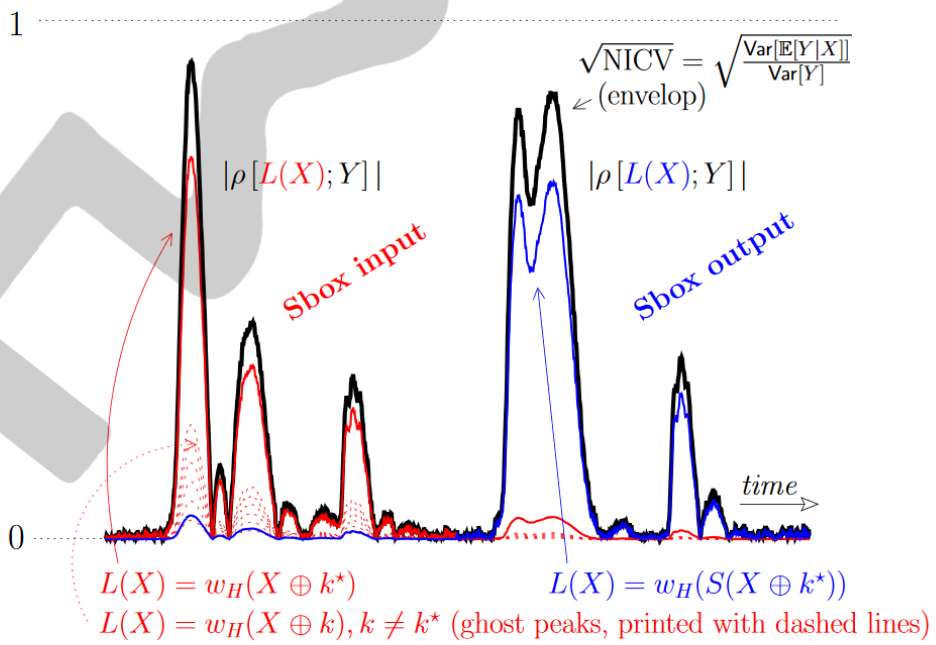
其中，

$$\rho^2[L(X);Y] = \rho^2[L(X);E[Y|X]] \cdot \rho^2[E[Y|X];Y]$$

等號左側即為相關係數能量攻擊(CPA)結果的平方。

由於 $\rho^2[L(X);E[Y|X]]$ 必然介於 0 與 1 之間，故 CPA 計算結果之絕對值 $\rho[L(X);Y]$勢必小於等於 $\rho[E[Y|X];Y]$的絕對值，也就是$\sqrt{NICV}$。等號成立於 $\rho^2[L(X);E[Y|X]] = 1$，即 L 達到 100%精確度，與能量曲線呈完全的線性關係之時。換句話說，當執行相關係數功率分析 (Correlation Power Analysis, CPA)時，若能使用更貼近事實的洩漏模型，則在正確金鑰下得到的數值越靠近 NICV，藉此，我們便可判斷模型的好壞，印證 NICV 特性的第四點所述。

圖五描述了 NICV，以及不同洩漏模型和能量曲線的相關係數間的關係。NICV 以粗體黑線表示。可以觀察到，在左側 Sbox 輸入的部分，好模型與能量曲線的相關係數(紅色實線)遠高於不優的模型(紅色虛線、藍色實線)；然而，藍色實線所代表的模型，其描述的對象其實是 Sbox 的輸出，因此在圖右側 Sbox 輸出



圖五、NICV 與不同模型的 CPA 數值

的部分，藍色實線具有相當高的相關係數，反之紅實線與紅虛線的相關係數趨近於 0。其實不難發現，對於一個洩漏模型而言，能量曲線的相關係數只會在相應的位置跳起，其餘不相干的位置則接近零位線；反之，NICV 在每個可能有資訊洩漏的位置都會跳起，並且數值都大於等於 CPA 所得的結果。

回顧定義。從定義中明顯可見，NICV 數值與金鑰無關，故計算 NICV 並不能取得金鑰資訊。NICV 數值也與洩漏模型無關，只需要能量曲線以及對應的明(密)文，即可求出 NICV 曲線，因此並不需要實際取得裝置，亦不須了解裝置的實作細節。這項特性的確讓我們節省不少初步檢測的成本，也使得將晶片安全性的評估提前至開發階段，甚至是規格制定階段都成為可能。

接著仔細審視 NICV 定義的內涵。定義中給定任一時間點 t，分母為「所有能量曲線在時間 t 上的資料點的變異數」，而分子為「分母的資料點依明文分組後，組間的變異數」，此內容與 變異數分析或變方分析（Analysis of variance, ANOVA）有關。NICV 這指標描述的是，一個裝置在加密不同中間值時，所產生的能量消耗是否可辨別？倘若答案是肯定的，那麼攻擊者很可能找到某一種手法，在攻擊這個時間點時取得破解金鑰的相關資訊。

假設N代表裝置運行中無可避免的電子雜訊。則模擬結果為:

$$Y = cL(X, K^*) + N$$

在時間 t 時，該裝置執行加密運算產生的功耗遵守 L 模型(洩漏模型)。於是，

$$Var\big[E[Y|X]\big] = E[(E[Y|X])^2] - (E[E[Y|X]])^2$$

$$= \sum p(x)\cdot E[Y|X=x]^2 - E[Y]^2 = \sum p(x)\cdot E[L(x)]^2 - E[L]^2$$

$$= Var[L]$$

而 $Var[Y] = Var[L(X, K^*)] + Var[N]$，因此，

$$NICV = \frac{Var[E[Y|X]]}{Var[Y]} = \frac{1}{\frac{1}{SNR}+1}$$

其中信噪比或訊號雜訊比(Signal-to-Noise Ratio, SNR)，在信號處理領域廣泛被使用，也是在 SCA 領域重要的指標之一。SNR 定義為目標信號變異數與雜訊變異數的比值，也就是說當 SNR 越大時，越容易從原始信號辨識出目標信號，雜訊造成的影響也越小。換言之，能量曲線上信噪比越高的資料點，其資訊洩漏程度也更嚴重，NICV 的數值自然越大。此結論亦可由上式直接獲得。

另外，值得一提的是，由於不考慮金鑰，NICV 顯然不同於能量模板的建立。相較之下，NICV 彈性得多，也能以相當省力的方法進行資料點的快速篩

選。在進入下一節之前，我們必須再次強調兩點:

- NICV 只負責找出「潛在」洩漏資訊的位置。並非所有 NICV 值偏高的地方都能執行有效的攻擊。
- 就攻擊方而言，NICV 扮演的角色是輔助道具，而非攻擊手法。

（二） TVLA (Test Vector Leakage Assessment)

延續上一節的 NICV，本節著眼於另一種旁通道資訊洩漏指標，稱為 TVLA。而這種指標分為非指定(non-specific)與指定(specific)兩類，本文提到的 TVLA 均為非指定的 TVLA，詳細相關資訊需參考其他文件。

下表列出本節與上節使用的符號

符號	涵義	備註
X	明文	1 位元組
k	正確金鑰	1 位元組
L或 $l(X,k)$	標準化的洩漏模型*	$E[L]=0$ 且 $Var(L)=1$。
$Y=\epsilon L+N$	實錄或仿真的能量曲線	ϵ為常數，$N\sim N(0,\sigma^2)$為電子雜訊

要計算 TVLA，我們首先以某個明文位元組 X 為篩選標準，將能量曲線分成兩組: F 組使用正確金鑰與該明文 X，而 R 組則用同一把正確金鑰與所有隨機明文 (廣義而言，F 組包含於 R 組之中)。TVLA 使用了 Welch's T-test，旨在檢定兩群體之間是否有所差異。

欲檢定 F 組與 R 組兩群體的期望值是否相等，我們以全體變異數來判別兩期望值間的距離是否過於遙遠。倘若兩者理論均值相等，卻由抽樣(記錄加密運算的功耗)得到相去甚遠的兩個均值，是件稀奇古怪的事，理論上發生的機率甚小，以至於我們寧可選擇相信兩群體期望值並不相等，意即加密過程當中必然洩漏了些攻擊者可利用的資訊。延續上一章所討論的，先前就已介紹過 TVLA 的概念。

TVLA 的計算公式如下:

$$TVLA=\frac{\mu_r-\mu_f}{\sqrt{\frac{{\sigma_r}^2}{n_r}+\frac{{\sigma_f}^2}{n_f}}}$$

其中 μ_r , σ_r , n_r 分別為 R 組能量曲線的均值、標準差與數量；F 組同理。
如同前段所述，假設檢定的虛無假設與對立假設分別為 $\mu_r=\mu_f$ 與 $\mu_r\neq\mu_f$。
通過檢定(即統計結果支持拒絕虛無假設)的條件為:

$$|TVLA|>4.5$$

此條件在自由度大於 1000 的情況下，具備 99.999%的信心水準，因為

$$\Pr\{|TVLA| > 4.5\} < 0.00001$$

由此可見，要成功拒絕虛無假設的條件相當嚴苛。很多時候，我們會重複檢定兩次(當然，兩次檢定的材料不同)。如果能量消耗曲線上的某一時刻兩次檢定都通過，那麼該時刻有資訊洩漏將具備 99.999%的信心水準。

（三） TVLA、NICV 與 SNR 的轉換公式

TVLA 可以寫為(此處不證明):

$$\widehat{TVLA}_x \xrightarrow[Q\to\infty]{} \sqrt{Q}\frac{E[Y|X=x]-E[Y]}{\sqrt{Var[Y|X=x]+Var[Y]}}$$

於是，定義 TVLA 的漸進常數為

$$TVLA_x = \frac{E[Y|X=x]-E[Y]}{\sqrt{Var[E[Y|X=x]]+Var[E[Y]]}}$$

又 $E[Y] = \epsilon E[L] = 0$，故

$$TVLA_x = \frac{\epsilon l(x,k)}{\sigma}$$

[NICV 與 SNR]

回顧 NICV 與 SNR 的定義，並以 ϵ 與 σ 表示。

$$NICV = \frac{Var[E[Y|X]]}{Var[Y]} = \frac{1}{1+\frac{\sigma^2}{\epsilon^2}}$$

其中，

$$Var\big[E[Y|X]\big] = Var[\epsilon L] = \epsilon^2$$
$$Var[Y] = Var[\epsilon L] + Var[N] = \epsilon^2 + \sigma^2$$

而

$$SNR = \frac{Var[E[Y|X]]}{E[Var[Y|X]]} = \frac{\epsilon^2}{\sigma^2}$$

因此，NICV 與 SNR 間的轉換公式為:

$$NICV = \frac{1}{1+\frac{1}{SNR}}$$

[TVLA 與 SNR]

$$SNR = Var[TVLA_X]$$

[說明]

$$\mathrm{Var}[TVLA_X] = Var\left[\frac{\epsilon L}{\sigma}\right] = \frac{\epsilon^2}{\sigma^2} = SNR$$

[TVLA 與 NICV]

由於定義限制，TVLA 只能將樣本分為兩組，因此我們先討論(不證明) NICV 也只分為同樣兩組的情況，符號表示為$NICV_2$。

$$NICV_2 = \frac{1}{\frac{1}{TVLA^2} + \frac{n}{C}({\sigma_1}^2 - {\sigma_2}^2)\left(\frac{1}{n_2} - \frac{1}{n_1}\right) + 1}$$

其中，$\mathrm{C} = ({\mu_1}^2 - {\mu_2}^2)^2$。

若是兩群樣本數相同，即 $n_1 = n_2 = \frac{n}{2}$，則上式可化簡為:

$$NICV_2 = \frac{1}{\frac{1}{TVLA^2} + 1}$$

之後，若想將 NICV 由兩組延伸至 k 組，則僅需依照以下公式即可:

$$NICV_k = \frac{k-1}{k}\sum_{i=1}^{k} NICV_2^i$$

其中，$NICV_2^i$ 表示對應到 TVLA 的 F 組為 k 組當中的第 i 組，而 R 組則為其他 $k-1$ 組的集合。

Reference

[1] W. Diffie and M. E. Hellman, “New directions in cryptography,” *IEEE Trans. Inf. Theory*, vol. 22, no. 6, pp. 644–654, Nov. 1976.

[2] [Online] Available: https://en.wikipedia.org/wiki/Public-key_cryptography

[3] P. W. Shor, “Algorithms for quantum computation: Discrete logarithms and factoring,” in *Proc. 35th Annu. Symp. Foundations of Computer Science*. Piscataway, NJ, USA, Nov. 1994, pp. 124–134.

[4] Y. Wang, Y. Li, Z. Yin, and B. Zeng, “16-qubit IBM universal quantum computer can be fully entangled,” Quantum Information, 4(1):46,2018.

[5] [Online] Available: https://en.wikipedia.org/wiki/Post-quantum_cryptography

[6] [Online] Available: https://csrc.nist.gov/news/2019/pqc-standardization-process-

2nd-round-candidates
[7] [Online] Available: https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Workshops-and-Timeline
[8] H. Wang, B. Sheng, and Q. Li, “Elliptic curve cryptography-based access control in sensor networks,” *International Journal of Sensor Networks*, 2006.
[9] C. Pendl, M. Pelnar, M. Hutter, "Elliptic curve cryptography on the WISP UHF RFID Tag," in *Proc. 7th Int. Workshop RFID Security*, pp. 32-47, Jun. 26–28, 2011.
[10] Z. Liu, X. Huang, Z. Hu, M. K. Khan, H. Seo, L. Zhou, "On emerging family of elliptic curves to secure internet of things: ECC comes of age," *IEEE Trans. Dependable Secure Comput.*, vol. 14, no. 3, pp. 237-248, May/Jun. 2017.
[11] [Online] Available: https://en.wikipedia.org/wiki/Side-channel_attack
[12] Paul C Kocher, "Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems," in *Proc. 16th Annual International Cryptology Conference (Advances in Cryptology - CRYPTO)*, Santa Barbara, CA, USA, Aug., 1996, pp. 104-113.
[13] P. C. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," in *Proc. 19th Annual International Cryptology Conference (Advances in Cryptology - CRYPTO)*, Santa Barbara, CA, USA, Aug., 1999, pp. 388-397.
[14] D. Agrawal, B. Archambeault, J. R. Rao, and P. Rohatgi, "The EM side-channel(s)," in *Proc. Conf. on Cryptographic Hardware and Embedded Systems (CHES)*, San Francisco Bay, CA, USA, Aug. 2002, pp. 29-45.
[15] D. Genkin, A. Shamir, and E. Tromer, “RSA Key Extraction via Low-Bandwidth Acoustic Cryptanalysis” in *Proc. 34th Annual International Cryptology Conference (Advances in Cryptology - CRYPTO)*, Santa Barbara, CA, USA, Aug., 2014, pp. 444-461.
[16] G. Camurati, S.Poeplau, M. Muench, T. Hayes, and A. Francillon, “Screaming channels: When electromagnetic side channels meet radio transceivers,” in *Proc. 25th ACM Conf. on Comput. and Commun. Security (CCS)*, Toronto, ON, CA, Oct. 2018, pp. 163-177.
[17] A. Park, K. -A. Shim, N. Koo, and D. –G. Han, “Side-Channel Attacks on Post-Quantum Signature Schemes based on Multivariate Quadratic Equations – Rainbow and UOV,” *IACR Trans. Crypt. Hardware Embed. Syst., Vol. 2018, Issue 3*, pp.500-523, 2018.
[18] H. Maghrebi, T. Portigliatti, and E. Prouff, “Breaking cryptographic implementations using deep learning techniques,” in *Proc. Int. Conf. Security Privacy Appl. Cryptography Eng.*, 2016, pp. 3–26.
[19] M. Carbone, V. Conin, M.-A. Cornélie, F. Dassance, G. Dufresne, C. Dumas, E. Prouff, and A. Venelli, “Deep Learning to Evaluate Secure RSA Implementations,”

IACR Trans. Crypt. Hardware Embed. Syst., *Vol. 2019*, *Issue 2*, pp.132-161, 2019.

[20] Matthias Kannwischer, "pqm4: Testing and Benchmarking NIST PQC on ARM Cortex-M4," *Conf. on Second PQC Standardization*, Santa Barbara, CA, USA, Aug., 2019

[21] [Online] Available: https://en.wikipedia.org/wiki/Normal_distribution

[22] [Online] Available: https://en.wikipedia.org/wiki/Gamma_distribution

[23] D. Wackerly, W. Mendenhall, R. Scheaffer, *Mathematical Statistics with Applications*, 7th edition, Thomson, 2008.

[24] 林其昌，基礎統計學於旁通道分析領域的應用

[25] S. Bhasin, J.-L. Danger, S. Guilley, and Z. Najm, "NICV: Normalized Inter-Class Variance for Detection of Side-Channel Leakage," *Cryptology ePrint Archive*, *Nov. 2013*.

[26] D. B. Roy, S. Bhasin, S. Guilley, A. Heuser, S. Patranabis, and D. Mukhopadhyay, "Leak Me If You Can: Does TVLA Reveal Success Rate?" *Cryptology ePrint Archive, Dec. 2016.*