Probabilidades de referencia para aplicar en la detección de Infraestructuras de Clave Pública anómalas

Antonio Castro Lecthaler, Marcelo Cipriano, Eduardo Malvacio

{antonio.castrolechtaler; cipriano1.618; edumalvacio}@gmail.com,

CriptoLab. Escuela Superior Técnica – Instituto Universitario del Ejército Argentino – Cabildo 15. A1406CCC – Ciudad Autónoma de Buenos Aires, Argentina

Abstract: Este trabajo permite calcular la probabilidad teórica de hallar primos repetidos en una muestra determinada de certificados digitales emitidos, libres de sesgos. Tales valores pueden servir de referencia para elaborar un procedimiento estadístico que permita auditar y fiscalizar el comportamiento de una Infraestructura de Clave Pública (PKI), y así poder detectar anomalías, si existieran, en su funcionamiento y evitar vulnerabilidades de esa naturaleza.

Keywords: PKI, RSA**,** Certificados Digitales.

1 Introducción

Los certificados emitidos por una Infraestructura de Clave Pública (PKI por sus siglas en inglés: Public Key Infrastructure) en entornos y sistemas militares como del ámbito civil, redes Públicas o Privadas, Lan´s, o Wan´s o asimismo Internet, tienen amplia difusión. Entre otras aplicaciones de los mismos se pueden destacar: logueo y autenticación de usuarios, equipos y sistemas, cifrado y firma digital, no repudio, determinación de claves de sesión, etc.

Los certificados que emite una PKI, incluyen entre otros, un módulo *m* y un número *e* (generalmente 65537) conocidos como *"clave pública"* y un número *d* llamado *"clave privada"*. El valor *m*, que tiene un tamaño *t* (medido en bits) se obtiene por el producto de 2 valores primos. Este trío *(m,e,d)* es calculado por la PKI al momento de solicitar el certificado digital correspondiente y entregado a un usuario que será su poseedor.

Se produce una vulnerabilidad[1] si la PKI manifiesta alguna anomalía al calcular los valores *m* o se emiten certificados en los que dos o más usuarios comparten algún factor primo de sus respectivos módulos.

[1] La seguridad del *Sistema RSA* se basa en la dificultad de factorizar en tiempo aceptable, módulos *m*, (por ejemplo, t=1024, 2048 o 4096 bits como los empleados en la actualidad) y por ello preservar la clave secreta *d*. Con el conocimiento de uno de los factores primos de un determinado módulo, se puede calcular de manera sencilla el otro factor y con él la clave *d*, previo paso, que se calcula sin dificultad.

Dicha información permite eludir la seguridad ofrecida por RSA y obtener sin dificultad la clave privada, pudiendo entonces acceder a la información que se pretende proteger.

La complejidad de los sistemas actuales es muy grande y la detección de determinado tipos de errores no es sencilla [7]. El método tradicional es la lectura y fiscalización de las líneas de código que forman la PKI. La detección de errores, por otro lado, es una realidad que tiene muchos antecedentes, dentro de los cuales se pueden consultar [1,9].

Una interesante discusión puede darse acerca de la naturaleza de dichos errores: inocentes "bugs" que superaron las pruebas y se filtraron para ser detectados años después de su creación o fueron "plantados" con la intención de debilitar la seguridad.

Otros investigadores [8] han evaluado más de un millón de certificados de clave pública y descubrieron que cerca del 5% de los mismos compartían factores primos. ¿Este es un valor esperable, dada la magnitud de la muestra analizada o está fuera de las posibilidades en vista del tamaño de los módulos analizados y la cantidad de primos posibles?

Este trabajo y sus antecedentes[2-6], determina la *Función de Probabilidad* de encontrar o no colisiones de factores primos sobre un *espacio muestral* variable, compuesto de certificados. Tales probabilidades (que se encuentran libres de anomalías y sesgos pues fueron calculados teóricamente) podrán utilizarse como valores de referencia, para auditar y fiscalizar el comportamiento de una PKI determinada.

En los puntos 2 y 3 se presenta un modelo probabilístico para la realización del experimento E, sus posibles resultados. La metodología para determinar los resultados antes mencionados y finalmente se obtendrán las fórmulas que permitan calcular las probabilidades en los que en muestras de tamaño *mu*, haya o no primos repetidos de tamaño *t*.

En el punto 4 se presenta la Función de Probabilidad que rige en una PKI teórica, libre de anomalías y sesgos.

En el punto 5 se presentan diferentes fórmulas para calcular factoriales grandes, pues las fórmulas obtenidas en los puntos anteriores los requieren.

En el punto 6 se presentan las conclusiones y posibles continuaciones de esta investigación: determinación del modelo probabilístico oculto en una PKI a analizar, empleando herramientas provenientes de la *inferencia estadística*. Y finalmente la comparación entre los valores teóricos y los obtenidos empíricamente, que permitirán determinar la presencia de sesgo o anomalías. El cierre final sería la elaboración de un software que permita auditar PKI.

2 Cantidad de certificados emitibles por una PKI

Sea la *hipótesis*[2] *H1: el tamaño de los valores primos generados por la PKI, llamado aquí b, es la mitad del valor del tamaño t de los módulos.* Por ejemplo, si *t=1024* entonces los valores primos serán de *b=512* bits de tamaño.

Sea P_1 el conjunto de números primos de tamaño *b*.

$$P_1 = \{p \,/\, p \text{ primo};\ 2^{b-1} < p < 2^b\}. \tag{1}$$

El cardinal o cantidad de elementos de P_1 -llamado aquí p_1 - puede calcularse con una fórmula asociada al Teorema de los Números Primos[3]:

$$p_1 = Card(P_1) \approx \pi(2^b) - \pi(2^{b-1}). \tag{2}$$

$$p_1 \approx \frac{2^b}{ln\, 2^b} - \frac{2^{b-1}}{ln\, 2^{b-1}} = \frac{2^{b-1}}{ln\, 2}\left(\frac{2}{b} - \frac{1}{b-1}\right). \tag{3}$$

Sea M_1 el conjunto de todos los módulos públicos que se pueden determinar a partir de los elementos del conjunto P_1:

$$M_1 = \{m/m = pq;\ p \neq q;\ p, q \in P_1\}. \tag{4}$$

Se asume aquí otra hipótesis de trabajo *H2: la PKI no emitirá certificados obtenidos como el producto del mismo factor primo*. Es decir que el módulo público no será un número cuadrado.

El cardinal de M_1 (que se indicará por m_1) es la cantidad de subconjuntos de 2 elementos del conjunto P_1, ya que cada módulo público es el producto de 2 valores primos y por la conmutatividad del producto, no importa el orden en el que se los multiplique.

$$m_1 = Card(M_1) = \binom{p_1}{2} = \frac{p_1(p_1 - 1)}{2}. \tag{5}$$

---

[2] A lo largo de este trabajo se irán asumiendo diferentes hipótesis respecto al contexto y entorno. Se irán describiendo y numerando medida que se vayan presentando.

[3] Conjeturado por el matemático alemán *Carl Gauss* (1777-1855) y demostrado de forma independiente por el matemático belga *Charles-Jean de la Vallée Poussin (1866-1962)* y el matemático francés *Jacques Hadamard* (1865-1963).

3. El experimento E, modelo probabilístico y su Función de Probabilidad

3.1 Definición del experimento E

Se propone la realización del experimento *E*:

Experimento E: *solicitar a la PKI la cantidad de* ***mu*** *certificados y con ellos formar el conjunto MU llamado "muestra".*

$$MU = \{m \,/\, m \text{ es un módulo público de tamaño } t\}. \tag{6}$$

$$Card(MU) = mu. \tag{7}$$

Se asume la *Hipótesis 3: la obtención de los módulos m está libre de sesgos*. Por lo que esto determina un modelo probabilístico en el que la probabilidad de obtener cualquier módulo es equiprobable.

3.2 Resultados del experimento

Este experimento puede tener 2 *resultados* o *eventos:*

$$R = \{r_1 ; r_2\}. \tag{8}$$

- r_1: que en el conjunto *MU* no haya módulos *m* que compartan algún factor primo. En cuyo caso se dirá que *no hay colisiones de factores primos*.
- r_2: que en el conjunto *MU* haya 2 o más módulos *m* que repitan factores primos. En cuyo caso se dirá que *sí hay colisiones de factores primos.*

Para determinar cuál es el resultado del experimento[4] se empleará el Máximo Común Divisor de todos los módulos, tomados de a pares:

$$\forall m_i, m_j \in MU (i \neq j); mcd(m_i, m_j) \in \{1; p\}, p \in P_1. \tag{9}$$

Por lo tanto, si todos los valores obtenidos del *mcd* es 1, el experimento tuvo resultado r_1 pues no se verifica que haya primos repetidos en dichos módulos. Caso contrario, el experimento tuvo resultado r_2.

[4] Se puede solicitar a la PKI los factores primos del módulo, junto con el resto de la información del certificado. Luego bastará con revisar si en la muestra hay o no primos repetidos. Ambas pruebas, el cálculo del mcd y como esta última descripta, tienen costo computacional que en el marco de este trabajo no se analizará. Lo recomendable es elegir la que tenga menor complejidad de las dos.

El proceso de aplicar el experimento *E* a todos conjuntos posibles *MU*, determina 2 conjuntos: R_1 y R_2:

$$R_1 = \{ MU / R(MU) = r_1 \}. \quad (10)$$

$$R_2 = \{ MU / R(MU) = r_2 \}. \quad (11)$$

Tales conjuntos tienen las siguientes propiedades:

a) Disjuntos:

$$R_1 \cap R_2 = \varnothing. \quad (12)$$

b) Complementarios:

$$R_1 \cup R_2 = EM(E). \quad (13)$$

Donde *EM(E)* es el *Espacio Muestra del experimento E*, es decir todos los conjuntos *MU* de tamaño *mu*.

$$EM(E) = m_1^{mu} \quad (14)$$

Sea la *hipótesis H4: la PKI no "recuerda" los certificados emitidos*. Por lo tanto podría repetir módulos en su espacio muestral.[5]

Definimos la Función de Probabilidad, de acuerdo a las propiedades de R_1 y R_2

$$p(R_1) + p(R_2) = 1. \quad (15)$$

$$p(R_2) = 1 - p(R_1). \quad (16)$$

Para determinar el valor de estas probabilidades utilizaremos la teoría clásica[6]. Para tal fin se deberá calcular el cardinal de cada conjunto y el total del espacio muestral.

3.3 Cardinalidad de R1 y R2

El cardinal de R_1 es la cantidad de conjuntos *MU*, de tamaño *mu*, formado por módulos de tamaño *t*, en los que se verifica que *no* haya colisiones de primos.

---

[5] En caso de asumir otra hipótesis respecto a la PKI, esto conllevaría el consecuente cambio de tamaño del espacio muestral.

[6] La Teoría de Probabilidades fue iniciada por *Pierre de Fermat (1601-1665)* y *Blaise Pascal (1623-1662)*. Sin embargo se debe *a Pierre-Simon Laplace (1749-1827)* la primera definición axiomática de probabilidad: sea la probabilidad de un suceso el cociente entre cantidad de resultados favorables del suceso y la cantidad total de resultados posibles.

Dados los conjuntos P_1 y M_1 indicados en las fórmulas (1, 3) respectivamente, se tiene que el 1er elemento de *MU* puede ser cualquiera de los m_1 elementos de M_1.

El segundo elemento, debiera ser un módulo coprimo con el primero de la muestra. Para tal fin determinamos al conjunto P_2 como el conjunto de números que resulta de quitar los números *p* y *q* que forman al primer elemento del conjunto P_1.

$$P_2 = P_1 - \{p; q\}. \tag{17}$$

$$p_2 = Card(P_2) = Card(P_1) - 2. \tag{18}$$

Sea M_2 el conjunto de todos los módulos que se puedan generar con P_2, cuyo cardinal m_2 se calcula como el combinatorio de todos los elementos de P_2, tomados de a 2.

$$M_2 = \{m/m = pq; p \neq q; p, q \in P_2\}. \tag{19}$$

$$m_2 = Card(M_2) = \binom{p_2}{2} = \frac{p_2(p_2 - 1)}{2}. \tag{20}$$

$$m_2 = \binom{p_1 - 2}{2} = \frac{(p_1 - 2)(p_1 - 3)}{2} \tag{21}$$

Con el mismo razonamiento, el 3er elemento de la muestra será un módulo coprimo con el primero y con el segundo elemento, tomados de a dos. Para ello se determina el conjunto de números primos P_3 que resulta de quitar de P_2 los factores *p* y *q* que determinan al segundo elemento.

$$P_3 = P_2 - \{p; q\}. \tag{22}$$

$$p_3 = Card(P_3) = Card(P_2) - 2 = Card(P_1.) - 4. \tag{23}$$

Sea M_3 el conjunto de todos los módulos que se puedan generar con P_3.

$$M_3 = \{m/m = pq; p \neq q; p, q \in P_3\}. \tag{24}$$

$$m_3 = Card(M_3) = \binom{p_3}{2} = \frac{p_3(p_3 - 1)}{2}. \tag{25}$$

$$m_3 = \binom{p_1 - 4}{2} = \frac{(p_1 - 4)(p_1 - 5)}{2}. \tag{26}$$

En general, para cualquier valor *i* entre *1* y *mu*, se tiene:

$$Pi = P_{i-1} - \{p; q\}. \tag{27}$$

$$p_i = Card(P_i) = Card(P_{i-1}) - 2 = Card(P_1.) - 2(i-1). \tag{28}$$

Sea M_i el conjunto de todos los módulos que se pueden obtener como producto de elementos de P_i:

$$M_i = \{m/m = pq;\ p \neq q;\ p,q \in P_i\} \tag{29}$$

$$m_i = Card(M_i) = \binom{p_i}{2} = \frac{p_i(p_i - 1)}{2}. \tag{30}$$

$$m_i = \binom{p_i - 2(i-1)}{2} = \frac{[p_1 - 2(i-1)][p_1 - 2(i-1) - 1]}{2} \tag{31}$$

.

Se puede observar como cada cardinal de los conjuntos M_i se van expresando en función del cardinal de M_1.

Siguiendo este procedimiento hasta llegar al último módulo de la muestra. El módulo número *mu* que es la cantidad prevista en el *experimento E*.

$$P_{mu} = P_{mu-1} - \{p; q\}. \tag{32}$$

$$p_{mu} = Card(P_{mu}) = Card(P_{mu-1}) - 2 = Card(P_1) - 2(mu - 1). \tag{33}$$

Sea M_{mu} el conjunto de todos los módulos que se puedan generar con P_{mu}

$$M_{mu} = \{m/m = pq;\ p \neq q;\ p,q \in P_{mu}\}. \tag{34}$$

$$m_{mu} = Card(M_{mu}) = \binom{p_{mu}}{2} = \frac{p_{mu}(p_{mu} - 1)}{2}. \tag{35}$$

$$m_{mu} = \binom{p_1 - 2(mu-1)}{2} = \frac{[p_1 - 2(mu-1)][p_1 - 2(mu-1) - 1]}{2}. \tag{36}$$

Todos los módulos del conjunto *MU*, son coprimos, tomados de a 2.

Finalmente, el cardinal del conjunto de todas las muestras de *mu* módulos, en las que no hay colisiones de primos, queda determinado por:

$$Card(R_1) = \prod_{i=1}^{mu} m_i = \prod_{i=0}^{mu-1} \binom{p_1 - 2i}{2}. \tag{37}$$

$$Card(R_1) = \frac{\prod_{i=0}^{2(mu-1)} (p_1 - i).}{2^{mu}}. \tag{38}$$

Luego,

$$Card(R_1) = \frac{p_1!}{2^{mu}(p_1 - 2(mu-2))!}. \tag{39}$$

3.4 Cardinalidadp de R2

Dado que R_1 y R_2 son disjuntos y complementarios tal como se ha mostrado en (12-14), se tiene que:

$$Card(R_2) = EM(E) - Card(R_1). \tag{40}$$

$$Card(R_2) = m_1^{mu} - \frac{p_1!}{2^{mu}(p_1 - 2(mu-2))!}. \tag{41}$$

Siendo m_1 la cantidad de módulos que se pueden calcular con los factores primos del conjunto P_1, cuyo cardinal es el valor p_1 y mu que es la cantidad de módulos en cada muestra.

3.5 Función de Probabilidad

Tal como se expresa en (16-17), entonces:

$$p(R_1) = \frac{card(R_1)}{m_1^{mu}}. \tag{42}$$

$$p(R_1) = \frac{\frac{p_1!}{2^{mu}[p_1 - 2(mu-2)]!}}{m_1^{mu}}. \tag{43}$$

Luego, por (5):

$$p(R_1) = \frac{\frac{p_1!}{2^{mu}[p_1 - 2(mu-2)]!}}{\frac{[p_1(p1-1)]^{mu}}{2^{mu}}}. \tag{44}$$

$$p(R_1) = \frac{p_1!}{[p_1 - 2(mu-2)]![p_1(p1-1)]^{mu}}. \tag{45}$$

$$p(R_2) = 1 - \frac{p_1!}{[p_1 - 2(mu-2)]![p_1(p1-1)]^{mu}}. \tag{46}$$

4 Cálculo de factoriales grandes

Estas fórmulas requieren resolver factoriales muy grandes, cuya complejidad computacional dificulta dicho cálculo. A modo de ejemplo, se darán algunas fórmulas para aproximar el valor de los factoriales:

$$n! = e^{\ln n!} \approx e^{n(\ln n - 1)}. \tag{47}$$

$$n! \approx n^n e^{-n} \sqrt{2\pi n}. \tag{48}$$

$$n! \approx n^n e^{-n} \sqrt{\pi} \sqrt[6]{8n^3 + 4n^2 + n + \frac{1}{30}}. \tag{49}$$

Conociéndose (47) y (48) como las fórmulas de *Stirling*[7] y (49) de *Ramanujan*[8].

$$n! \approx \sqrt{2\pi} \left(\frac{n + \frac{1}{2}}{e} \right)^{n+\frac{1}{2}}. \tag{50}$$

$$n! \approx n^n e^{-n} \sqrt{\pi} \sqrt{2n + \frac{1}{3}}. \tag{51}$$

$$n! \approx n^n e^{-n} \sqrt{2\pi \left(n + \frac{1}{6} + \frac{1}{72n} - \frac{31}{6480n^2} - \frac{139}{155520n^3} + \frac{9871}{6531840n^4} \right)}. \tag{52}$$

Conocidas estas expresiones como las fórmulas de *Burnside*[9], *Gosper*[10] y *Batir*[11], respectivamente.

5 Conclusiones y futuros trabajos

Se han presentado las fórmulas para calcular las probabilidades matemáticas de hallar en una muestra colisiones de primos, calculados a partir de una fuente libre de sesgos y anomalías.

Se puede asumir como hipótesis la existencia de *permanencia estadística*, es decir que a lo largo de la realización del experimento E, aplicado a una PKI determinada, su modelo probabilístico desconocido puede ser revelado a través de herramientas estadísticas.

---

[7] *James Stirling* (1692-1770). Matemático escocés.

[8] *Srinivasa Ramanujan* (1887-1920). Matemático indio. No dejó una demostración de su fórmula. Fue demostrada en 2000 por la matemática rusa *Ekatherina Karatsuba*.

[9] *William Burnside* (1852-1927). Matemático inglés.

[10] *Ralph Gosper, Jr.* (1943-). Matemático y científico de computadoras estadounidense.

[11] *Necdet Batir* (1959 -). Matemático turco.

Habría anomalía en el comportamiento de la PKI si se verifican divergencias entre los valores de referencia propuestos en este trabajo y los obtenidos por "experiencia directa"

Quedan así planteadas cuestiones para continuar la investigación y poder elaborar un software que pueda auditar y fiscalizar anomalías en Infraestructuras de Clave Pública.

6 Referencias

[1] Bello L, Bertacchini M. *"Generador de Números Pseudo-Aleatorios Predecible en Debian"*. III Encuentro Internacional de Seguridad Informática. Manizales, Colombia. Octubre 2009.

[2] Benaben, A; Castro Lechtaler, A; Cipriano, M; Foti, A. *"Development, testing and performance evaluation of factoring algorithms whit additional information*" XXVIII Conferencia Internacional de la Sociedad Chilena de Computación. Santiago de Chile. 2009.

[3] Castro Lechtaler, C; Cipriano, M; Benaben A; Quiroga, P. *"Study on the effectiveness and efficiency of an algorithm to factorize N given e and d*". IX Seminario Iberoamericano en Seguridad de las Tecnologías de la Información, La Habana, CUBA. 2009.

[4] Castro Lechtaler, A; Cipriano, M. "*Detección de anomalías en Oráculos tipo OpenSSL por medio del análisis de probabilidades*". XVII Congreso Argen-tino de Ciencias de la Computación CACIC 2011. La Plata, Buenos Aires, Octubre 2011.

[5] Castro Lechtaler, Antonio, Cipriano Marcelo; Malvacio Eduardo; Cañón, Sebastián; *Procedure for the Detection of Anomalies in Public Key Infrastruture (RSA Systems)*. XIII Simposio Argentino de Tecnología, 41 Jornadas Argentinas de Informática e Investigación Operativa JAIIO – SADIO. La Plata, Buenos Aires, Agosto 2012.

[6] Castro Lechtaler, Antonio; Cipriano, Marcelo; Malvacio, Eduardo. *Experimental detection of anomalies in public key infrastructure*. XVIII Congreso Argentino de Ciencias de la Computación CACIC 2012. Bahía Blanca, Buenos Aires, Octubre 2011.

[7] Glass, Robert *"Facts and Fallacies of Software Engineering*". Addison-Wesley Professional, 2003

[8] Lenstra, A; Hughes, J; Augier, M y otros. Ron was wrong, Whit is right. e-print International Association for Cryptologic Research. 15 Feb 2012. http://eprint.iacr.org/2012/064.

[9] Young A and Yung M. *An Elliptic Curve Asymmetric Backdoor in Open-SSL RSA Key Generation.* Chapter 10. Cryptovirology. 2006.