Procedimiento para la detección empírica de Infraestructuras de Clave Pública anómalas o inseguras

Antonio Castro Lecthaler, Marcelo Cipriano, Eduardo Malvacio

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

Criptolab. Escuela Superior Técnica - Facultad de Ingeniería - Universidad Nacional de la Defensa (UNDEF) - Cabildo 15. A1406CCC – Ciudad Autónoma de Buenos Aires, Argentina

Abstract: El presente trabajo ofrece un procedimiento empírico susceptible de ser codificado para la creación de un software Auditor de una Infraestructura de Clave Pública (PKI). El mismo consiste en el cálculo de los parámetros estadísticos producidos por una PKI teórica – y por lo tanto libre de errores y defectos- y la obtención de los mismos parámetros de una PKI real. Así se podrá realizar el análisis comparativo de los mismos y determinar si dicho sistema se ajusta o no al comportamiento esperado, detectando sistemas defectuosos o inseguros.

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

1 Introducción

Muchos son los usos y aplicaciones de la **Criptografía de Clave Pública** o **Asimétrica**. En ella podemos hallar el Sistema **RSA** que posibilita "confidencialidad" y "autenticación" a través de su esquema de cifrado y firma digital, respectivamente.
Para poder llevar adelante estos servicios se puede montar una **Infraestructura de Clave Pública** (**PKI** por sus siglas en inglés) en los servicios de redes y demás. Estos sistemas ¿podrían contener errores que permitan su vulnerabilidad? Confiados en su "buen funcionamiento" los usuarios podrían tener comprometida su seguridad y ser blanco fácil de ataques.
Para poder detectar comportamientos "no aceptables" de estos sistemas, el presente trabajo y sus antecedentes presentan un ***Procedimiento Probabilístico-Estadístico para la Determinación Experimental de Infraestructuras de Clave Públicas Anómalas***. El mismo analiza el comportamiento de las **PKI´s** de acuerdo a la distribución de los factores primos que cada "*certificado digital*" contiene en su "*módulo público*" *m*. El mismo es factible de ser codificado y poder así crear un **Software Auditor**.
En los parágrafos 2 y 3 se tratará acerca de los conceptos generales de **la Criptografía de Clave Pública**, el sistema **RSA,** generalidades acerca de las **PKI.** También se presentarán algunas de los últimos descubrimientos acerca de debilidades e inseguridad de los mismos.
En el parágrafo 4 se presenta la fórmula que permite calcular la *Función de Probabilidad* de no hallar/hallar colisiones o repeticiones de factores primos en muestras de certificados digitales. Dado que la misma requiere el cálculo de

factoriales muy grandes, se presentan también fórmulas de aproximación de los mismos que no tengan la misma complejidad computacional del procedimiento original.
En el parágrafo 5 se presenta, finalmente, el procedimiento empírico para la detección de PKI, junto con las fórmulas que permiten calcular los valores de referencia probabilísticos y su proceso estadístico provenientes de la inferencia estadística.

2 Criptografía de Clave Pública.

2.1 Orígenes

Durante miles de años, conocidos como *La Era de la Criptografía Clásica*, sólo existía la Criptografía de *Clave Privada* o *Simétrica*: la misma clave, que se empleaba para cifrar un mensaje, debía usarse para descifrarlo. La clave debía ser compartida por el emisor y el receptor.
Esto, sin embargo, traía consigo un ineludible escollo: el "*Problema del Intercambio de Claves*" (PIC). En algún momento ambas entidades tuvieron que acordar la clave a utilizar. ¿cómo acordar la clave sin que la misma esté comprometida por usar un canal inseguro?
En el año 1976 **Whitfield Diffie** y **Martin Hellman** publican un trabajo en el que resuelven el PIC. Proponen lo que hoy se conoce como "**el Intercambio de Claves de Diffie-Hellman**" (DH) [1].
En 1977 **Ronald Rivest, Adi Shamir** y **Leonard Adleman**[1] publican un memo técnico ante el **Massachusetts Institute of Technology (MIT)**. En dicho trabajo presentan un sistema que permite cifrar/descifrar un mensaje eludiendo el PIC. A su vez el mismo principio permite la realización de un esquema de "*Firma Digital*", adicionando el servicio de "*Autenticación*" al ya ofrecido servicio de "*Confidencialidad*". Al año siguiente, 1978, publican el artículo por el que se conoce mundialmente al **Sistema RSA** [2].
La **Asociación de los Sistemas Informáticos ACM** (por sus siglas en inglés: **Association for Computing Machinery**), fundada en 1947, ofrece anualmente el máximo galardón en esta disciplina: el Premio Turing. Por sus importantes aportes a la criptografía en el año 2002 fue otorgado a Rivest, Shamir y **Adleman**. Y recientemente, en el año 2015, los galardonados fueron Diffie y Hellman [3].
Los trabajos fundacionales de *DH* y *RSA* dieron nacimiento a lo que se conoce como la "*Criptografía de Clave Pública*" (CCP) o "*Criptografía Asimétrica*". Este tipo de Criptografía utiliza *Aritmética Modular*. Dentro de esta área de la Matemática, se abordan dos problemas sin solución hasta hoy[2].

[1] No se usa el orden alfabético de sus apellidos pues así lo quisieron sus autores. El sistema que propusieron fue bautizado RSA por las siglas de sus apellidos y en ese orden.

[2] El "Problema del Logaritmo Discreto" (PLD) para el esquema propuesto por Diffie y Hellman y el "Problema de la Factorización de Números" (PFN) para el esquema de Rivest, Shamir y Adleman.

2.2 Vulnerabilidades descubiertas recientemente.

Desde su nacimiento la Criptografía de Clave Pública ha sido blanco de ataques y la comunidad científica ha buscado e investigado sus vulnerabilidades. Hasta el presente, no se han hecho avances matemáticos significativos en ninguno de los problemas matemáticos que se encuentran detrás de los esquemas criptográficos mencionados. Y desde ese punto de vista, se puede considerar a la CCP como segura.
Muchos de los problemas y debilidades hallados corresponden al nivel de la implementación. Entre las más recientes, en el año 2015, varias vulnerabilidades, conocidas como **LogJam** [4], han sido documentadas para DH[3].
Lo mismo ocurre para el esquema **RSA**[5]. Debilidades a nivel hardware [6, 7] como a nivel software. En particular cabe mencionar la investigación llevada adelante por el **Dr. Lenstra** [8] el cual, junto a otros investigadores han evaluado más de un millón de certificados de clave pública y descubrieron que cerca del 5% de los mismos compartían factores primos. Esta cantidad de certificados, que tienen comprometida su seguridad ¿es un valor alcanzable 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?

3 Infraestructura de Clave Pública.

3.1 Qué es una PKI.

En entornos y sistemas, tanto militares como del ámbito civil en redes Públicas o Privadas, Lan´s, o Wan´s como asimismo Internet, tienen amplia difusión las **Infraestructuras de Clave Pública** (**PKI** por sus siglas en inglés: **Public Key Infrastructure**). Una **PKI** entrega, a sus usuarios, "*certificados*". Entre otras aplicaciones, con ellos se pueden realizar logueos 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**[4], 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.
Una vulnerabilidad[5] interna o propia de la **PKI** se produce si la misma manifiesta alguna anomalía al calcular los valores *m*.

[3] Entre los consejos que los investigadores sugieren para resolver parte de este problema, se propone generar un grupo **Diffie-Hellman** de 2048 bits.

[4] El X.509 es un estándar de la **Unión Internacional de Telecomunicaciones** (**ITU: International Telecommunication Union**, por sus siglas en inglés) que estipula, entre otras cosas, los tipos y el formato de los certificados, el algoritmo para validarlos, etc.

[5] 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

El conocimiento de esta vulnerabilidad permitiría a un "atacante" eludir la seguridad ofrecida por el algoritmo **RSA** y obtener sin dificultad las claves privadas comprometidas, pudiendo entonces acceder a la información que se pretende proteger o suplantar identidades de usuarios, etc.

3.2 Por qué se producen las vulnerabilidades.

Los sistemas actuales tienen un alto nivel de complejidad y detectar determinados tipos de errores no es fácil [9]. La lectura y fiscalización de todas las líneas de código que forman la **PKI** puede resultar ser una tarea ardua.
La detección de errores que provocan disminución de la seguridad en estos sistemas es una realidad con muchos antecedentes. Por ejemplo una vulnerabilidad expuesta en la emisión de certificados digitales en **OpenSSL** de **Debian** [10].
Estos "errores" ¿son sólo 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?

4 Cálculo de la Función de Probabilidad de encontrar o no colisiones de factores primos.

4.1Obtención de la fórmula.

Se presenta la fórmula que permite calcular la probabilidad que 2 o más certificados digitales tengan primos compartidos, dentro de una muestra de tamaño *mu*. Una version más detallada de la obtención de esta función puede encontrarse en [11].
Sea *t* el tamaño, medido en bits, de los módulos públicos *m*. Sea *b* el tamaño en bits de los factores primos de *m*. Por ejemplo, si *t=1024* entonces los valores primos tendrán el tamaño *b=512* bits.
Sea P el conjunto de números primos de tamaño *b*.

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

El cardinal o cantidad de elementos de *P* -llamado aquí *p* - puede calcularse con una fórmula asociada al *Teorema de los Números Primos*[6]:

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

---

determinado módulo, se puede calcular de manera sencilla el otro factor y con él la clave *d*, previo paso, que se calcula sin dificultad.

[6] 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).

$$p \approx \frac{2^{b}}{\ln 2^{b}} - \frac{2^{b-1}}{\ln 2^{b-1}} = \frac{2^{b-1}}{\ln 2}\left(\frac{2}{b} - \frac{1}{b-1}\right). \quad (3)$$

Sea *M* el conjunto de todos los módulos públicos que se pueden determinar a partir de los elementos del conjunto *P*:

$$M = \{m / m = pq; p \neq q; p, q \in P\}. \quad (4)$$

El cardinal de *M* (que se indicará por *m*) es la cantidad de subconjuntos de 2 elementos del conjunto P, 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.

$$m = Card(M) = \binom{p}{2} = \frac{p\ (p-1)}{2}. \quad (5)$$

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

$$Card(R) = \prod_{i=1}^{mu} m_i = \prod_{i=0}^{mu-1} \binom{p-2i}{2}. \quad (6)$$

$$Card(R) = \frac{\prod_{i=0}^{2(mu-1)} (p-i).}{2^{mu}}. \quad (7)$$

Luego

$$Card(R) = \frac{p!}{2^{mu}(p-2(mu-2))!}. \quad (8)$$

Para calcular la **Función de Probabilidad** se aplicará la definición clásica de probabilidad de **Laplace**[7]. Se asume como hipótesis que la **PKI** no almacena ni registra los primos que ya ha usado. Se tiene entonces:

$$p(R) = \frac{card(R)}{m^{mu}}. \quad (9)$$

$$p(R) = \frac{\frac{p!}{2^{mu}[p-2(mu-2)]!}}{m^{mu}}. \quad (10)$$

Luego, operando convenientemente:

[7] Definición laplaciana de probabilidad: casos favorables sobre casos totales.

$$p(R) = \frac{p!}{[p - 2(mu-2)]![p\ (p1-1)]^{mu}}. \tag{11}$$

Y su probabilidad complementaria sería:

$$p(\overline{R}) = 1 - \frac{p!}{[p - 2(mu-2)]![p\ (p-1)]^{mu}}. \tag{12}$$

4.2 Cálculo de factoriales grandes

A continuación se exponen algunas fórmulas para aproximar el valor de factoriales grandes, como los que se usan en el cálculo de la función de probabilidad de las fórmulas (11) y (12). La complejidad computacional de la fórmula tradicional para la obtención de factoriales es demasiado elevada para que las fórmulas sean viables. Es por ello que la introducción de fórmulas de aproximación puede justificarse.

$$n! = e^{ln\,n!} \approx e^{n(ln-1)}. \tag{13}$$

$$n! \approx n^n e^{-n} \sqrt{2\pi n}. \tag{14}$$

$$n! \approx n^n e^{-n} \sqrt{\pi} \sqrt[6]{8n^3 + 4n^2 + n + \frac{1}{30}}. \tag{15}$$

Las fórmulas (13) y (14) se conocen como las fórmulas de **Stirling**[8] y la fórmula (15) de **Ramanujan**[9].

$$n! \approx \sqrt{2\pi} \left(\frac{n + \frac{1}{2}}{e} \right)^{n+\frac{1}{2}}. \tag{16}$$

$$n! \approx n^n e^{-n} \sqrt{\pi} \sqrt{2n + \frac{1}{3}}. \tag{17}$$

$$n! \approx n^n e^{-n} \sqrt{2\pi \left(n + \frac{1}{6} + \frac{1}{72n} - \frac{31}{6480n^2} - \frac{139}{155520n^3} + \frac{9871}{6531840n^4} \right)}. \tag{18}$$

Estas últimas 3 expresiones son conocidas como las fórmulas de **Burnside**[10], **Gosper**[11] y **Batir**[12], respectivamente.

---

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

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

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

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

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

5 Algoritmo auditor de PKI´s

5.1 Valores de referencia de la distribución binomial de una PKI libre de sesgos

La fórmula (11) permite calcular la probabilidad que en una muestra de certificados digitales de tamaño *mu* hayan al menos 2 de ellos que compartan un valor primo.
Habrá muestras con primos repetidos y en otras no. Esto representa un **Experimento de Bernoulli**. Esto es: ensayos o experimentos aleatorios e independientes entre sí, con dos resultados posibles y complementarios, llamados "*éxitos*" y "*fracasos*".
Como consecuencia, se obtiene la conocida **Distribución Binomial:**

$$X \sim B(n, p(R)). \tag{19}$$

$$f(x) = \binom{n}{x} p(R)^x [1 - p(R)]^{n-x}. \tag{20}$$

Donde *x* es variable aleatoria discreta que representará la cantidad de éxitos/fracasos[13] que se pueden hallar en *n* ensayos o muestras de tamaño *mu.* Y *p(R)* es la probabilidad de no hallar/hallar colisiones de primos en una muestra de tamaño *mu*, tal como indica la fórmula (11).
Algunos de los parámetros de esta distribución de probabilidad son la *media* y la *varianza* (con la que se puede computar el *desvío estándar*, que es la raíz cuadrada de la *varianza*):

$$\bar{x} = np(R). \tag{21}$$

$$s^2 = np(R)[1 - p(R)]. \tag{22}$$

5.2 Detección experimental de Infraestructuras de Clave Pública Anómalas

Finalmente hemos arribado al objetivo principal de este trabajo y su línea de investigación portadora: la detección experimental de Infraestructuras de Clave Públicas anómalas por medio de su comportamiento empírico. Tal proceso es susceptible de ser automatizado. Así poder crear un "*algoritmo auditor de PKI´s"*.
Se acepta como hipótesis del trabajo la existencia de "*permanencia estadística*": el procedimiento experimental al que se someterá la **PKI** permitirá revelar su comportamiento desconocido a través de herramientas estadísticas.
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*"

[13] Se puede optar por la probabilidad que en la muestra no haya primos repetidos o su complemento. Ambas calculadas en (11) y (12) respectivamente.

El procedimiento probabilístico-estadístico es el siguiente:

1- Determinación de los parámetros de evaluación de la **PKI:** tamaño *m* de los módulos y por consiguiente de *b*, el tamaño de sus factores primos.
2- Cálculo del valor *p*, según la fórmula (3).
3- Determinación del tamaño de cada muestra *mu*.
4- Cálculo[14] del valor *p(R)*, según fórmula (11).
5- Cálculo de los parámetros, según fórmulas (21) y (22).
6- Solicitar a la **PKI** *certificados digitales* de manera que con ellos se los puedan agrupar en *n* muestras de tamaño *mu*.
7- Contar en cuántas de tales muestras no hay/hay primos repetidos (búsqueda de colisiones).
8- Con los valores obtenidos en el paso anterior, calcular los parámetros empíricos de la media y la varianza.
9- Comparar los valores teóricos predichos en el punto 5 con los obtenidos en el punto 8.
10- Determinar si la **PKI** manifiesta un comportamiento anómalo o no, de acuerdo a la comparación realizada en el punto 9.

Para poder evaluar si hay o no primos repetidos en una muestra hay varias alternativas. Si la **PKI** puede indicar los valores primos que utilizó en el cálculo del módulo público de cada certificado. Si no se puede tener información de los primos, se podría recurrir al procedimiento indicado en [12].

6 Conclusiones y futuros trabajos

Se han presentado las fórmulas que permiten calcular el comportamiento libre de sesgos y vulnerabilidades de una **PKI** ideal. Se ha presentado un procedimiento probabilístico-estadístico que permite determinar empíricamente el comportamiento real de una **PKI**.
La comparación entre ambos valores (el predicho teóricamente y el verificado experimentalmente) permite determiner el comportamiento anómalo o no de una determinada **Infraestructura de Clave Pública.**
Este proceso puede ser llevado a algún format de código y así crea un "**Software Auditor de Infraestructuras de Clave Públicas**".
Futuros trabajos de investigación podrían probar el procedimiento presentado, como así también resolver algunos interrogantes que no se han abordado aún, entre otros:

- las fórmulas de aproximación para el cálculo de grandes factoriales ¿introducirán incertidumbres que interfieran en la determinación de las anomalías a detectar?

[14] Se podrá optar por una de las fórmulas de aproximación de factoriales grandes, tal como se expone en el parágrafo 4.2.

- ¿cuál es la mejor de ellas para aplicar en el procedimiento, dada la relación costo computacional/beneficio?
- ¿cuál es la brecha aceptable entre los parámetros teóricos y los empíricos?

7 Agradecimientos

El presente trabajo es parte de un **Proyecto de Desarrollo Tecnológico y Social (PDTS)** [13] de la **Escuela Superior Técnica -Facultad de Ingeniería-** perteneciente a la **Universidad Nacional de la Defensa (UNDEF).**
El proyecto tiene como entidades adoptantes al **Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF)** y al **Consejo Profesional de Ingeniería de Telecomunicaciones, Electrónica y Computación (COPITEC).**
A ellos el agradecimiento de los autores.

8 Referencias

1. Diffie, W. y M.E.Hellman. "New directions in cryptography", IEEE Transactions on Information Theory 22 pp. 644-654. 1976.
2. R. Rivest, A. Shamir, L. Adleman. “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems.” Communications of the ACM, Vol. 21 (2), pp.120–126. 1978.
3. http://amturing.acm.org/byyear.cfm consultada el 20/6/2016.
4. Adrian, D.; Bhargavan, K.; Durumeric, Z.; Gaudry, P.; Green, M.; Halderman, J.; Heninger, N.; Springall, D.; Thomé, E.; Valenta, L.; VanderSloot, B.; Wustrow, E.; Zanella-Béguelin, S.; Zimmermann, P. “Imperfect Forward Secrecy: How Diffie-Hellman Fails in Practice.” Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Pages 5-17. ACM New York, NY, USA. 2015.
5. Boneh, D. Twenty Years of Attacks on the RSA Cryptosystem. Notices of the American Mathematicals Society, Volume 46, Number 2. Providence, 1999.
6. Chen, S.; Wang, R.; Wang, X.; Zhang, K. “Side-Channel Leaks in Web Applications: a Reality Today, a Challenge Tomorrow”. IEEE Symposium on Security & Privacy. Oakland, 2010.
7. Pellegrini, A.; Bertacco, V.; Austin, T. Fault-based attack of RSA authentication. Proceedings Design, Automation & Test in Europe Conference & Exhibition. The IEEE Council. Dresden, 2010.
8. Lenstra, A; Hughes, J; Augier, M y otros. Ron was wrong, Whit is right. e-print International Association for Cryptologic Research. 2012. http://eprint.iacr.org/2012/064.
9. Glass, Robert “Facts and Fallacies of Software Engineering”. Addison-Wesley Professional, 2003.
10. Bello L, Bertacchini M. “Generador de Números Pseudo-Aleatorios Predecible en Debian”. III Encuentro Internacional de Seguridad Informática. Manizales, Colombia. Octubre 2009.
11. Castro Lechtaler, Antonio; Cipriano, Marcelo; Malvacio, Eduardo. “Detección de Infraestructuras de Clave Pública anómalas”. XXI Congreso Argentino de Ciencias de la Computación CACIC 2015. Junín, Buenos Aires. Octubre 2015.
12. Cipriano, M. “Factorización de N: recuperación de factores primos a partir de las claves pública y privada”. XIV Congreso Argentino de Ciencias de la Computación. CACIC 2008. Chilecito, La Rioja. Octubre 2008.
13. http://www.iese.edu.ar/investigacion.html#antecedentes consultada el 20/6/2016.