PENENTUAN DISTRIBUSI DARI BANYAKNYA '*HIT*' KERANDOMAN BARISAN BILANGAN BINER PADA METODE *OVERLAPPING TEMPLATE MATCHING TEST*

Sarini Abdullah[1], Dheni Triadi Sudewo[2]

[1]*Departemen Matematika FMIPA UI, Depok*
[2]*Alumni Departemen Matematika FMIPA UI, Depok*

Email korespondensi : sarini@ui.ac.id,dheni.triadi@ui.ac.id

Abstrak

Pada penelitian ini dibahas mengenai penentuan distribusi dari banyaknya '*hit*' kerandoman barisan bilangan biner pada metode *Overlapping Template Matching Test*. Metode ini merupakan suatu metode yang terfokus pada sering atau tidaknya muncul 'pola' acak pada tiap blok barisan bilangan biner dengan menggunakan suatu *template*. Penentuan distribusi ini dimulai dengan menggunakan distribusi *Compound Poisson* , lebih khusus lagi menggunakan distribusi *Geometric Poisson*. Lebih lanjut lagi digunakan transformasi *Confluent Hypergeometric Function* (*Kummer's Function*). Selain itu, juga diberikan ilustrasi dalam menguji kerandoman barisan bilangan biner dengan menggunakan metode *Overlapping Template Matching Test*.

Kata kunci: distribusi, banyaknya hit, barisan bilangan biner, *Overlapping Template Matching Test*, *Compound Poisson*

PENDAHULUAN

Dalam bidang kriptografi, kerahasiaan suatu informasi merupakan hal yang paling penting. Terdapat banyak metode yang dapat digunakan untuk merahasiakan atau menghasilkan informasi tersebut. Salah satu contohnya adalah metode dalam menghasilkan bilangan acak. Terdapat dua tipe pembangkit yang dapat digunakan untuk menghasilkan bilangan yang bersifat acak: *Random Number Generator* (RNG) dan *Pseudorandom Number Generator* (PRNG).

Untuk mengetahui suatu barisan bilangan biner acak atau tidak, perlu dilakukan suatu pengujian. Untuk menguji kerandoman barisan yang dihasilkan oleh metode RNG maupun metode PRNG tersebut terdapat beberapa metode yang dapat digunakan. Salah satu metode tersebut adalah *Overlapping Template Matching Test*.

Pada metode *Overlapping Template Matching Test*, suatu barisan biner dengan panjang n akan dibagi sebanyak N blok dengan panjang masing - masing blok adalah M. Dari masing - masing blok inilah akan diuji apakah ada atau tidak suatu 'pola' yang menjadi acuan , yang disebut *template* B dengan panjang m. Berdasarkan *template* inilah akan diuji kerandoman barisan bilangan biner dengan cara melihat intensitas sering atau tidaknya muncul pola. Apabila terjadi kecocokan (misal *template* yang digunakan 11, dan bagian barisan bilangan yang dievaluasi adalah 11), maka untuk selanjutnya kejadian tersebut disebut '*hit*' (Rukhin, et al., 2001).

METODOLOGI

Penelitian ini merupakan suatu kajian teoritis berdasarkan studi literatur. Dengan asumsi bahwa barisan bilangan biner yang dihasilkan adalah random, maka akan ditentukan distribusi dari barisan bilangan tersebut. Pada penelitian ini, barisan bilangan biner, pembagian jumlah serta panjang

blok, dan *template* pengujian pola acak sudah diberikan. Sedangkan t*emplate* B yang digunakan adalah

$B = 111...111$ sepanjang m (*runs of ones*).

Mengacu pada asumsi bahwa barisan bilangan biner yang dihasilkan oleh suatu PRNG adalah random, maka didefinisikan suatu peubah acak indikator, I, dari setiap bilangan biner. Kemudian menggunakan distribusi Bernoulli, maka dapat dibuat suatu fungsi distribusi dari peubah acak I. Langkah selanjutnya adalah evaluasi banyaknya *hit* pada setiap blok dan penentuan formulasi untuk parameter yang diperlukan. Hal ini dilakukan dengan menggunakan sifat dan keterkaitan antara distribusi Bernoulli, Binomial, dan Poisson.

Menggunakan definisi dari distribusi *Compound Poisson*, maka akan ditunjukkan bahwa dari pendefinisian peubah acak I akan diperoleh suatu peubah acak yang memenuhi sifat dari distribusi *Compound Poisson*, dengan bentuk khusus adalah *Geometric Poisson*. Selanjutnya, dengan menggunakan *Confluent Hypergeometric Function* diperoleh perumusan akhir untuk fungsi kepadatan distribusi dari *Geometric Poisson*.

Notasi yang digunakan adalah sebagai berikut:ε menyatakan barisan bilangan biner, n adalah panjang dari ε, M adalah panjang *bit* tiap blok yang dibentuk dari ε, N merupakan banyaknya blok yang dibentuk dari ε, dengan $M \times N \leq n$, B adalah template sepanjang *m-bit.*

HASIL DAN PEMBAHASAN

Metode *Overlapping Template Matching Test* ini adalah salah satu metode yang digunakan untuk melihat pola pada barisan bilangan biner. Pada metode ini, pola yang dimaksud adalah kemunculan 11 ...111 sepanjang m kali pada barisan bilangan biner, untuk selanjutnya akan disebut dengan '*hit*' jika memiliki kecocokan (kesamaan) dengan *template* acuan yang digunakan.

Sebagai contoh, jika barisan bilangan biner yang diberikan adalah 110111101101110, dengan *template* yang digunakan adalah 11, maka pada barisan tersebut terjadi 7 *hit*. Hal ini diperoleh dengan melakukan pencocokan pola *template* 11 dengan bilangan biner pada barisan yang diberikan, dimulai dari angka awal pada barisan. Pencocokan dilakukan pada dua angka awal (yaitu 11, cocok dengan *template*, maka terjadi 1 *hit*), kemudian bergeser satu angka (yaitu 10; tidak cocok dengan *template*, *hit* yang terjadi masih 1), demikian seterusnya hingga angka terakhir pada barisan. Cara kerja inilah yang disebut dengan *overlapping template*, dimana baik terjadi *hit* maupun tidak, pencocokan berikutnya selalu hanya bergeser satu posisi ke angka selanjutnya.

Untuk suatu barisan bilangan biner ε dengan panjang n, yaitu:

$$\varepsilon = \varepsilon_1 \varepsilon_2 \cdots \varepsilon_M \varepsilon_{M+1} \cdots \varepsilon_{2M} \varepsilon_{2M+1} \cdots \varepsilon_{2M+1} \cdots \varepsilon_{2M} \cdots \varepsilon_{n-1} \varepsilon_n$$

dimana

$$\varepsilon_i = \begin{cases} 1, \text{ dengan probabilitas } p \\ 0, \text{ dengan probabilitas } q, \text{ dimana } q = 1 - p \end{cases}$$

ε_i adalah percobaan Bernoulli dengan ε_i dikatakan sukses jika $\varepsilon_i = 1$. ε adalah barisan dari Bernoulli *trials* karena merupakan percobaan Bernoulli yang dilakukan berkali-kali dan saling bebas. Lebih lanjut, pada setiap *trial* probabilitas suksesnya adalah sama yaitu p, sehingga dapat dikatakan bahwa:

$\varepsilon_i \sim \text{Be}(p), 1 \leq i \leq n$

Kemudian definisikan I_i untuk blok dengan panjang M dan *template* B dengan panjang m yaitu:

$$I_i = \prod_{j=i}^{i+m-1} \varepsilon_j, \quad 1 \leq i \leq M - m + 1$$

dimana

$$I_i = \begin{cases} 1, \text{jika terjadi 'hit'} \\ 0, \text{jika tidak terjadi 'hit'.} \end{cases}$$

Dapat didefinisikan W untuk blok dengan panjang M dan *template* B dengan panjang m pada blok ke - j yaitu:

$$W = \sum_{i=1}^{M-m+1} I_i = I_1 + I_2 + \ldots + I_{M-m+1}$$

untuk $i = 1, 2, \ldots, M - m + 1$.

Definisi 1. Untuk suatu $S = X_1 + X_2 + \ldots + X_N$ dengan

1. N merupakan peubah acak dari distribusi Poisson
2. $X_1, X_2, \ldots X_N$ adalah independen dan berdistribusi identik
3. N independen dengan $X_1, X_2, \ldots X_N$

Maka S dikatakan berdistribusi Compound Poisson (Kaas, R. et al, 2002).

Berdasarkan Nuel (2006), harus ada satu tambahan sifat khusus dalam distribusi *Compound Poisson*nya dengan variabel acak I_i merupakan varibel acak diskrit, yaitu

4. $P(I_i = k) = \frac{\lambda_k}{\lambda}, \lambda_k > 0$, dan $\sum_{k=0}^{1} \lambda_k = \lambda, \forall k \epsilon \mathbb{N}^*$

Sehingga pada pembahasan berikut akan ditunjukkan bahwa peubah acak W memenuhi keempat sifat di atas.

1. $N^* = M - m + 1$ merupakan variabel acak berdistribusi Poisson

   Berdasarkan definisinya $N^* = M - m + 1$ merupakan banyaknya pemeriksaan dengan *template* ukuran m dalam suatu blok dengan panjang M. Berdasarkan sifat dari distribusi Poisson, jelas bahwa $N^* = M - m + 1$ merupakan variabel acak yang berdistribusi Poisson.

2. I_i merupakan variabel acak yang memenuhi asumsi identik independen untuk $i = 1, 2, \ldots, M - m + 1$.

   a. Identik

$$E(I_i) = E\left(\prod_{j=i}^{i+m-1} \varepsilon_j\right) = E(\varepsilon_i.\varepsilon_{i+1}.\ldots\varepsilon_{i+m-1})$$

$$= E(\varepsilon_i).E(\varepsilon_{i+1}).\ldots E(\varepsilon_{i+m-1}), = p^m \text{, untuk } 1 \leq i \leq M-m+1$$

Maka $I_i \sim \text{Be}(p^m)$ untuk setiap $i = 1,2,\ldots,M-m+1$ sedemikian sehingga dapat dikatakan I_i identik.

b. Independen

$$\Pr(I_k = \text{'hit'} | I_j = \text{'hit'})$$

$$= \Pr(\varepsilon_k \ldots \varepsilon_{j+m-1} \ldots \varepsilon_{k+m-1} = 1 | \varepsilon_j \ldots \varepsilon_k \ldots \varepsilon_{j+m-1} = 1) = \left(\frac{1}{2}\right)^{k-j}$$

$$\Pr(I_k = \text{'hit'} | I_j = \text{tidak 'hit'})$$

$$= \Pr(\varepsilon_k \ldots \varepsilon_{j+m-1} \ldots \varepsilon_{k+m-1} = 1 | \varepsilon_j \ldots \varepsilon_k \ldots \varepsilon_{j+m-1} = 0) = \left(\frac{1}{2}\right)^{k-j}$$

Dengan cara serupa dapat ditunjukkan bahwa

$$\Pr(I_k = \text{tidak 'hit'} \mid I_j = \text{'hit'}) = \Pr(I_k = \text{tidak 'hit'} \mid I_j = \text{tidak 'hit'})$$

$$= \frac{1 - \left(\left(\frac{1}{2}\right)^{k-j}\right)}{\left(\frac{1}{2}\right)^m}$$

Sehingga dapat disimpulkan bahwa variabel acak I_i berdistribusi identik dan independen.

3. Variabel acak $N^* = M - m + 1$ independen terhadap I_i.

Jelas bahwa variabel acak $N^* = M - m + 1$ independen terhadap I_i. Berdasarkan definisi keduanya yaitu $N^* = M - m + 1$ merupakan banyaknya pemeriksaan yang dilakukan dalam suatu blok sedangkan I_i menyatakan '*hit*' atau tidaknya bagian dari blok tersebut yang sedang diperiksa, terlihat jelas bahwa keduanya tidak saling mempengaruhi, sehingga dikatakan $N^* = M - m + 1$ independen terhadap I_i.

4. $P(I_i = k) = \frac{\lambda_k}{\lambda}$, $\lambda_k > 0$, dan $\sum_{k=0}^{1} \lambda_k = \lambda, \forall k \epsilon \mathbb{N}^*$

Karena telah dibuktikan bahwa I_i berdistribusi Bernoulli, maka dapat dituliskan:

$$P(I_i = k) = (p^m)^k (1-p^m)^k, \quad k = 0,1$$

Pilih $\lambda_k = \lambda(p^m)^k(1-p^m)^k > 0$,dan $\sum_{k=0}^{1}\lambda_k = \lambda_0 + \lambda_1 = \lambda(p^m)^1 + \lambda(1-p^m)^1 = \lambda$.

Sehingga:

$$P(I_i = k) = (p^m)^k(1-p^m)^k = \frac{\lambda_k}{\lambda}, \lambda_k > 0 \text{ dan } \sum_{k=0}^{1}\lambda_k = \lambda \quad k = 0, 1$$

Sehingga terbukti bahwa W berdistribusi *Compound Poisson.* Untuk selanjutnya akan dituliskan $W \sim CP(\lambda_k)$.

Selanjutnya akan dilakukan penurunan distribusi dari W dengan cara menurunkan p.d.f dari W. Telah dibuktikan bahwa W berdistribusi *Compound Poisson*, dinotasikan $W \sim CP(\lambda_k)$, dengan parameter λ_k, dengan $\lambda_k > 0$ dan $\sum_{k=0}^{1}\lambda_k = \lambda$, λ_k adalah parameter untuk setiap I_i, yang menyatakan 'bobot' bahwa $I_1, I_2, \ldots, I_m$ masuk ke kelas, dimana:

$$W = \sum_{i=1}^{N^*} I_i = I_1 + I_2 + \ldots + I_{N^*}, \text{ dimana } N^* = M - m + 1$$

dengan N^* berdistribusi Poisson dan independen terhadap I_i, dimana I_i berdistribusi identik dan independen, dan

$$P(I_i = k) = \frac{\lambda_k}{\lambda}, \lambda_k > 0, \text{ dan } \sum_{k=0}^{1}\lambda_k = \lambda, \forall k \in \mathbb{N}^*$$

Lemma 1.

Jika $N \sim CP((\lambda_k)_{k\in\mathbb{N}^*})$ dengan $\sum_{k=1}^{\infty}\lambda_k = \lambda$ maka $\forall n \in \mathbb{N}^*$

$$P(N = n) = \sum_{m=1}^{n} e^{-\lambda}\frac{\lambda^m}{m!}\sum_{k_1,\ldots k_m\in\mathbb{N}^*} I_{\{k_1+k_2+\ldots+k_m=n\}}\frac{\lambda_{k_1}\times\ldots\times\lambda_{k_m}}{\lambda^m}$$

dan $P(N = 0) = e^{-\lambda}$.

(Nuel, 2006)

Karena $W \sim CP(\lambda_k)$, dengan $\sum_{k=0}^{1}\lambda_k = \lambda$, maka berdasarkan Lemma 1, $\forall k \in \mathbb{N}^*$,

$$P(W = i) = \sum_{l=1}^{i} e^{-\lambda}\frac{\lambda^l}{l!}\sum_{k_1,\ldots,k_l\in\mathbb{N}^*} I_{\{k_1+\ldots+k_l=i\}}\frac{\lambda_{k_1}\times\ldots\times\lambda_{k_l}}{\lambda^l}$$

dan $P(W = 0) = e^{-\lambda}$.

Kemudian akan dibuktikan bahwa W memiliki bentuk distribusi khusus dari distribusi *Compound Poisson*. Perhatikan bentuk dari W,

$$W = \sum_{i=1}^{M-m+1} I_i,$$

Pertama - tama perhatikan barisan bilangan biner dalam satu blok:

$$\varepsilon_i \sim \text{Be}(p) \text{ untuk } 0 \leq p \leq 1$$

Karena selanjutnya akan diturunkan distribusi dari W dibawah asumsi bahwa barisan bilangan biner yang diberikan adalah acak, maka $p = \frac{1}{2}$. Untuk *template* B dengan panjang m, probabilitas akan terjadi '*hit*' adalah $\frac{1}{2^m}$, dengan kata lain $p = \frac{1}{2^m}$. Banyaknya '*hit*' dalam satu blok tertentu berdistribusi binomial, dengan parameter $p = \frac{1}{2^m}$ adalah probabilitas terjadinya '*hit*' dan $M - m + 1$ adalah banyaknya pengecekan. Lebih lanjut distribusi dari banyaknya '*hit*' ini dapat didekati dengan distribusi Poisson, dengan parameter $\lambda = (M - m + 1)\frac{1}{2^m}$.

Jika $W \sim CP(\lambda_k)$, dengan $\lambda_k = \lambda(\theta)^k(1-\theta)^k$, $k = 0,1$, λ sebagai parameter untuk bagian Poisson, dan θ sebagai parameter untuk bagian Geometrik, maka dikatakan W memiliki suatu distribusi khusus dari *Compound Poisson* yaitu W berdistribusi *Geometric Poisson* (*Pólya-Aeppli*). Dinotasikan dengan $W \sim GP(\lambda, \theta)$.

Karena $W \sim GP(\lambda, \theta)$, maka untuk setiap $i \epsilon \mathbb{N}^+$

$$P(W = i) = \sum_{l=1}^{i} e^{-\lambda}\frac{\lambda^l}{l!}(1-\theta)^{i-l}(\theta)^l\binom{i-1}{l-1}$$

dimana $\binom{i}{l}$ merupakan koefisien binomial, dan $P(W = 0) = e^{-\lambda}$.

Berdasarkan Johnson, Kotz, & Kemp, 1996, dimana apabila W berdistribusi *Pólya-Aeppli*, dengan parameter λ sebagai parameter untuk bagian Poisson, dan parameter p sebagai parameter untuk bagian Geometriknya, maka dapat dituliskan:

$$W \sim \text{Poisson}(\lambda) \bigvee \text{Geometrik}(p)$$

dimana nilai parameter λ didefinisikan sebagai $\lambda = \frac{\eta}{p}$.

Kemudian akan didefinisikan suatu $q = 1 - p$, dengan $p = \frac{1}{2}$. Dengan $\lambda = \frac{\eta}{p}$ maka nilai $\eta = p\lambda = \frac{\lambda}{2}$. Sehingga berdasarkan Johnson, Kotz, & Kemp, 1996, bentuk distribusi dari W yaitu:

$$P(W = 0) = e^{-\eta}$$

$$P(W = i) = e^{-\eta}p^i\sum_{l=1}^{i}\binom{i-1}{l-1}\frac{\left(\frac{\eta q}{p}\right)^l}{l!}$$

dengan

$p = \frac{1}{2}$, maka $q = \frac{1}{2}$.

Sehingga untuk $\eta = \frac{\lambda}{2}$ bentuk distribusinya dapat ditulis:

$$P(W = i) = e^{-\eta}\left(\frac{1}{2}\right)^i\sum_{l=1}^{i}\binom{i-1}{l-1}\frac{\left(\frac{\eta \cdot 1}{2/2}\right)^l}{l!} = \frac{e^{-\eta}}{2^i}\sum_{l=1}^{i}\binom{i-1}{l-1}\frac{\eta^l}{l!}$$

Berdasarkan:

$$\sum_{l=1}^{i} \binom{i-1}{l-1} \frac{\eta^{l-1}}{l!} = \Phi(-i+1,2,-\eta)$$

Dengan

$$\Phi(-i+1,2,-\eta) = e^{-\eta}\Phi(i+1,2,\eta),$$

maka didapat bentuk distribusi dari W dengan menggunakan *Confluent Hypergeometric Function* (*Kummer's M*), yaitu:

$$\pi_i = P(W=i) = \frac{e^{-\eta}}{2^i} \sum_{l=1}^{i} \binom{i-1}{l-1} \frac{\eta^l}{l!}$$

$$= \frac{\eta e^{-2\eta}}{2^i} \Phi(i+1,2,\eta)$$

dengan $\Phi(a,b,z) =$ *Confluent Hypergeometric Function (Kummer's)* , (Abramowitz,M & Stegun,I.A. 1972).

KESIMPULAN

Penentuan distribusi dari banyaknya '*hit*' kerandoman barisan bilangan biner pada metode *Overlapping Template Matching Test* didapat melalui beberapa tahapan yaitu, pertama suatu barisan bilangan biner ε dengan panjang n, dibagi menjadi N blok dengan panjang masing - masing blok adalah M dengan syarat $n \geq MN$. Kemudian gunakan *template* B dengan panjang m untuk menguji tiap - tiap blok.

Dengan mendefinisikan I_i sebagai variabel acak yang menyatakan '*hit*' atau tidaknya bagian dari barisan suatu blok dengan menggunakan *template* B , dan W untuk suatu blok yang merepresentasikan banyaknya '*hit*' pada blok tersebut diperoleh bahwa W berdistribusi Geometric Poisson dengan pdf

$$\pi_i = P(W=i) = \frac{e^{-\eta}}{2^i} \sum_{l=1}^{i} \binom{i-1}{l-1} \frac{\eta^l}{l!} = \frac{\eta e^{-2\eta}}{2^i} \Phi(i+1,2,\eta)$$

DAFTAR PUSTAKA

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Washington, D.C.: Wiley-Interscience Publication.

Johnson, N. L., Kotz, S., & Kemp, A. W. (1996). *Univariate Discrete Distributions.* New York: Wiley-Interscience Publication, 2nd ed.

Kaas, R., Govaerts, M., Dhaene, J., & Denuit, M. (2002). *Modern Actuarial Risk Theory.* Boston/ Dordrecht/ London: Kluwer Academic Publishers.

Nuel, G. (2006). *Cummulative Distribution Function of a Geometric Poisson Distribution.* Paris: Laboratoire Statistique & Génome Press.

Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., et al. (2001). *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications.* Springfield: National Institute of Standards and Technology (NIST).