TEK BOYUTLU BİR KAOTİK FONKSİYONUN GERÇEK RASTSAL SAYI ÜRETİMİNE UYGUNLUĞUNUN İNCELENMESİ
A FEASIBILITY STUDY OF A 1D CHAOTIC MAP FOR TRUE RANDOM NUMBER GENERATION

İhsan Çiçek[1,2]*, Ali Emre Pusane*[2]*, Günhan Dündar*[2]

[1]TÜBİTAK BİLGEM UEKAE, Kocaeli
[2]Elektrik Elektronik Mühendisliği Bölümü
Boğaziçi Üniversitesi, İstanbul

ihsancicek@uekae.tubitak.gov.tr, ali.pusane@boun.edu.tr, dundar@boun.edu.tr

ÖZETÇE

Kriptografinin hayati ve vazgeçilmez ilkeli olan gerçek rastsal sayı üreteçleri bağımsız, eş dağılımlı ve tahmin edilemez sayılar üreten sistemlerdir. Bu çalışmada yapıca karmaşık, tasarımı ve analizi zor olan sürekli zaman kaotik rastsal sayı üreteçlerine alternatif oluşturan tek boyutlu bir kaotik fonksiyonun rastsal sayı üretimine uygunluğu lojistik fonksiyonundan yola çıkılarak geliştirilen bir matematiksel model üzerinden incelenmiş ve herhangi bir art-işlemciye gereksinim olmaksızın rastsal sayı üreteci olarak kullanılma potansiyeli değerlendirilmiştir. MATLAB ortamında gerçeklenen matematiksel model ile rastsal sayılar üretilerek NIST 800.22 istatistiksel testleri uygulanmıştır. Elde edilen sonuçlar tek boyutlu ayrık zaman kaotik fonksiyonların gerekli şartların sağlanması koşuluyla gerçek rastsal sayı üreteci olarak kullanılabileceğini ortaya koymuştur.

ABSTRACT

True Random Number Generators, vital and indispensable primitives of cryptography, are systems that generate identically distributed, independent, and unpredictable numbers. In this work, we study and evaluate the true random number generation potential and feasibility of a one dimensional chaotic function over a customly developed mathematical model of logistic map in MATLAB as an alternative example to existing topologically complex, hard to analyze and design continuous time chaos based true random number generators. We have generated random numbers using the model and applied NIST 800.22 statistical tests. Results reveal the true random number generation potential of one dimensional chaotic function when certain conditions are met.

1. GİRİŞ

Gerçek rastsal sayı üreteçleri (GRSÜ) kriptografik sistemlerin en önemli yapıtaşını oluşturur. Doğası gereği kriptografik bir sistemde hiçbir belirlenimci blok çıkışında girişindekinden daha fazla entropi üretemez. Bu nedenle sistemin entropisi ve tahmin edilemezliği temelde rastsal sayı üretecininkiyle (RSÜ) sınırlı olur. Tipik bir gerçek rastsal sayı üreteci Şekil 1'de gösterildiği gibi üç temel yapıtaşından oluşur: Entropi kaynağı, örnekleyici ve art-işlemci.

Şekil 1: Evrensel rastsal sayı üreteci.

Entropi kaynağı rastsallığın temel kaynağını oluşturur. Uygulamada direnç tabanlı bir devre elemanının ürettiği ısıl gürültü [1], ters kutuplanmış bir diyotun ürettiği çığ gürültüsü ya da osilatör seğirtimi [2] gibi belirlenimci yöntemlerle tahmin edilmesi imkansız olan fiziksel olaylara dayanan bileşenler entropi kaynağı olarak kullanılırlar.

Örnekleyici ise entropi kaynağından edinilen ayrık analog örnekleri sayısala çeviren birimdir ve entropi kaynağının istatisiksel özellikleri göz önüne alınarak tasarlanıp, karşılaştırıcı ve/veya flip floplar kullanılarak gerçeklenir. Gerçek entropi kaynakları üretim toleransları ya da hataları nedeniyle ideal olmayan istatistiksel özelliklere sahip olurlar. Bu durum üretilen bit katarının sıfır/bir oranında kendini ideal değer olan 0.5'ten sapma olarak gösterir.

Art-işlemci birimi, temelde bu istatistiksel sapmaları gidermek amacıyla tasarlanan matematiksel bir fonksiyon ya da algoritmadır. Yaygın şekilde kullanılan ardışıl iki bitin XOR işleminden geçirilmesi ve sıfır-bir ya da bir-sıfır geçişlerine göre sayı üreten Von Neumann algoritması örnek verilebilir. Art-işlemci birimi kullanılarak elde edilen istatistiksel iyileştirmeye karşı sayı üretim hızında düşme olur.

Kaos, başlangıç koşullarına hassas bağımlılık ve aperiyodik salınım üretebilme özelliğiyle GRSÜ uygulamalarında kendine yer bulur. Pozitif $Lyapunov$ üsteline sahip olan ka-

Bu çalışmayı destekleyen TÜBİTAK BİLGEM UEKAE'ye teşekkür ederiz.

otik bir sistemde başlangıç koşullarındaki en küçük bir hata zamanda evrilen sistemde büyük sapmalara neden olur. Kelebek etkisi olarak da bilinen bu özellik kaotik sistemlerin tahmin edilebilirliğini imkansızlaştırır. Matematiksel olarak belirlenimci olsa da uygulamada, kaos tabanlı bir GRSÜ'nün tahmin edilemezliği başlangıç koşullarının ortamda var olan ısıl gürültü nedeniyle sürekli değişmesine ve bu koşulları sonsuz bir kesinlikle ölçmenin imkansızlığına dayanır.

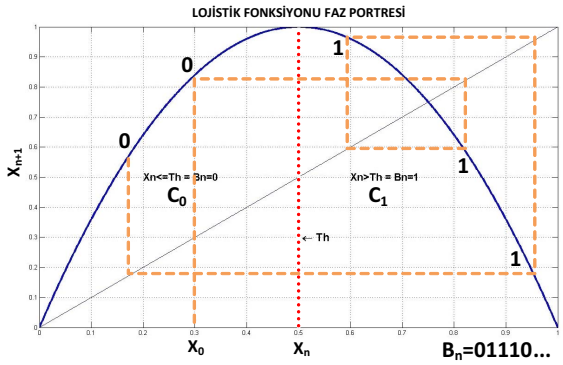
Bu çalışmada yapıca karmaşık, tasarımı ve analizi zor olan sürekli zaman kaotik GRSÜ'lerine alternatif oluşturan ayrık zaman tek boyutlu kaotik fonksiyonların rastsal sayı üretimine uygunluğu lojistik fonksiyonu örnek alınarak geliştirilen bir matematiksel model üzerinden incelenmiş ve herhangi bir artişlemciye ihtiyaç duymaksızın rastsal sayı üreteci olarak kullanılma potansiyeli değerlendirilmiştir. MATLAB ortamında gerçeklenen GRSÜ modeli ile rastsal sayılar üretilerek NIST 800.22 istatistiksel testleri uygulanmıştır. Elde edilen sonuçlar tek boyutlu ayrık zaman kaotik lojistik fonksiyonunun gerekli şartların sağlanması koşuluyla GRSÜ olarak kullanılabileceğini ortaya koymuştur.

2. GERÇEK RASTSAL SAYI ÜRETECİ MODELİ

Literatüre 19.yy'da P. F. Verlhulst tarafından tanıtılan lojistik fonksiyonu biyolojide av-avcı popülasyon dinamiğini modellemede kullanılır [3, 4]. Ayrık zamanlı lojistik fonksiyonu,

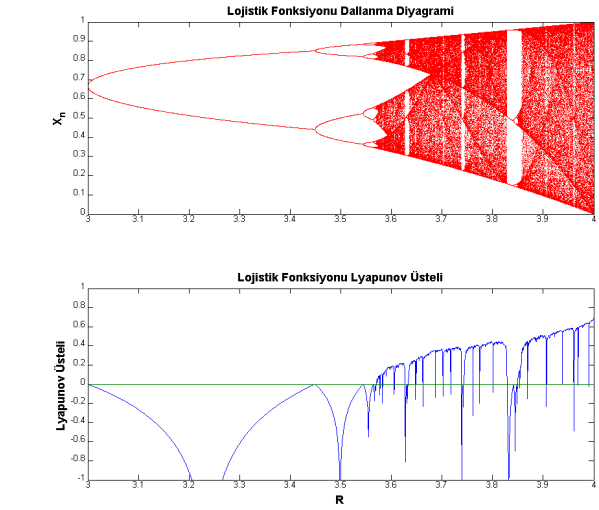
$$x(n+1) = Rx(n)(1-x(n)) \tag{1}$$

tanıma göre zamanda Şekil 2'de gösterildiği gibi evrilir.



Şekil 2: Lojistik fonksiyonundan rastsal sayı üretimi.

Lojistik fonksiyonunda R kaos kontrol parametresidir ve aldığı değere göre dinamik sistemin davranışı Şekil 3'teki dallanma grafiğinde gösterildiği gibi çeşitlilik gösterir. Pozitif $Lyapunov$ üsteli dinamik sistemin kaotik rejimde çalıştığının en belirgin göstergesidir ve en yüksek değerini ($\lambda = 0.6931$) $R = 4$'te alır. Kaotik sistemin üreteceği entropinin $Lyapunov$ üsteliyle orantılı olduğu gözönüne alınırsa en büyük $Lyapunov$ üsteline denk gelen R değerinin rastsal sayı üretimi için seçilmesi uygun olacaktır. Lojistik fonksiyonundan rastsal sayı üretmek için faz portresini birbiriyle örtüşmeyen parçalara bölünüp, ayrık zamanlı kaotik sistemin zamanda evrilirken takip ettiği yörüngenin faz portresindeki konumu kullanılabilir. Faz portresinin Şekil 2'de gösterilen $C_0 = [0, T_h)$ ve $C_1 = [T_h, 1]$ gibi iki üreteç bölgesinden oluştuğu varsayımı altında, kaotik yörüngenin faz portresindeki konumu ve T_h bölme parametresi kullanılarak,

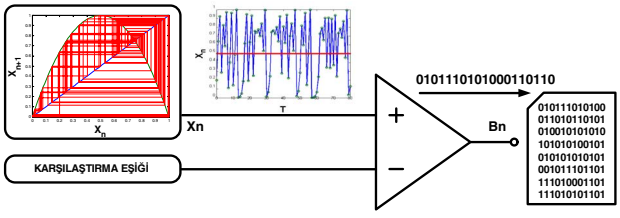


Şekil 3: Lojistik fonksiyonunun dallanma ve $Lyapunov$ üsteli grafiği.

$$S(x(n)) = \begin{cases} 0 & x(n) \in [0, T_h) \\ 1 & x(n) \in [T_h, 1] \end{cases} \tag{2}$$

uyarınca rastsal sayılar üretilebilir. Şekil 2'de rastsal seçilen $x(0) = 0.30$ ilk değerinden başlanarak ilk beş zaman adımı için rastsal sayı üretimi örnek olarak gösterilmiştir.

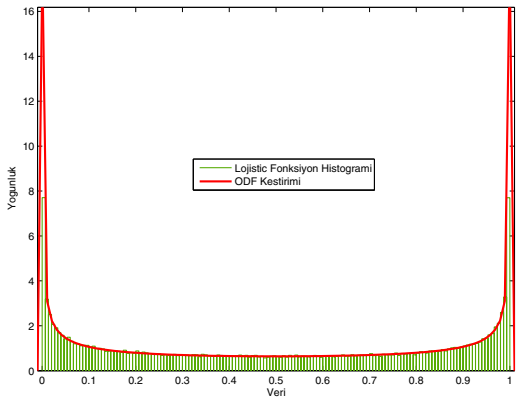
Şekil 4'te gösterilen lojistik fonksiyonunun entropi kaynağı olarak kullanıldığı GRSÜ MATLAB modelinde örnekleme, kaotik sistemin zamanda ilerleme adımı hızında gerçekleşir. MATLAB GRSÜ modeli rastsal bir ilk değerden evrilerek sayı üretir ve bu sayılar herhangi bir art-işlem uygulanmaksızın istatistiksel testler için ikili tabanda bir dosyaya kaydedilir.



Şekil 4: Lojistik fonksiyonu rastsal sayı üreteci modeli.

3. ENTROPİ KAYNAĞININ İSTATİSTİKSEL ANALİZİ

Entropi kaynağından rastsallığın nasıl örnekleneceğini belirlemek için öncelikle entropi kaynağının istatistiksel özelliklerinin anlaşılması gerekir. Kaotik sistemler ergodik oldukları için uzun vadeli istatistiksel özellikleri başlangıç koşullarından bağımsızdır. Entropi kaynağı olarak seçilen lojistik fonksiyonunun altında yatan olasılık dağılım fonksiyonunu belirlemek amacıyla MATLAB ortamında kodlanan GRSÜ modeli, rastsal seçilen ilk durumdan başlayarak yeterli veri üretecek şekilde (yüzbin bit) çalıştırılmış, elde edilen verinin normalize histogramından yola çıkılarak deneysel olasılık dağılımı numerik olarak Şekil 5'teki gibi elde edilmiştir.



Şekil 5: $R = 4$ için olasılık yoğunluk fonksiyonu kestirimi.

Elde edilen dağılımın beta dağılımına olan benzerliği nedeniyle en büyük olabilirlik kestirim yöntemiyle dağılıma ait şekil parametreleri $\alpha = 0.5, \beta = 0.5$, olarak hesaplanmıştır Şekil ve değer kümesi parametreleri $p = 0, q = 1$,

$$f_x(x) = \frac{(x-p)^{\alpha-1}(q-x)^{\beta-1}}{(q-p)^{\alpha+\beta-1}\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt}, \quad x \in [p,q], \quad \alpha,\beta > 0 \tag{3}$$

biçiminde verilen genel beta olasılık dağılım fonksiyonuna yerleştirilerek analitik olasılık dağılım fonksiyonu,

$$f_x(x) = \frac{1}{\pi\sqrt{x(1-x)}} \tag{4}$$

olarak elde edilmiştir. Numerik analizden yola çıkılarak elde edilen olasılık dağılım fonksiyonunun, teorik analizle türetilen değişimsiz ölçüte (invariant measure) eşit olduğu doğrulanmıştır [5]. Birikimli olasılık dağılım fonksiyonu, (4)'te verilen lojistik fonksiyonuna ait olan olasılık dağılım fonksiyonunun integrali alınarak,

$$F_x(x) = \int_{-\infty}^{T_h} f_x(x)dx = \frac{2}{\pi} sin^{-1}\sqrt{T_h} \tag{5}$$

biçiminde elde edilmiştir.

4. EŞ OLASILIKLI VE BAĞIMSIZ ÖRNEKLEME

İdeal bir GRSÜ'de üretilen her sayının eş olasılığa sahip olması ve ardışıl olarak üretilmiş iki sayı arasında istatistiksel bağımsızlık bulunması gerekir. Entropi kaynağından alınan örnekler bir bitlik analog-sayısal çevirici gibi davranan karşılaştırıcı tarafından nicemlenerek sayısal bitlere dönüştürülür. GRSÜ modelinde üretilen sayıların istatistiksel özelliklerine etkiyen iki kontrol parametresi vardır: lojistik fonksiyonunun kaotik davranışını kontrol eden R parametresi ve faz portresini ikiye bölen eşik değeri parametresi, T_h. Kaotik sistemden en yüksek entropi seviyesinde örnekler alabilmek için $R = 4$ seçilmiştir. GRSÜ modeli için sistemin sıfır üretme olasılığı, (4)'teki olasılık dağılım fonksiyonu kullanılarak,

$$\begin{aligned} P(0) = P(x(n) < T_h) &= \int_{-\infty}^{T_h} f_x(x)dx \\ &= \frac{2}{\pi} sin^{-1}\sqrt{T_h} \end{aligned} \tag{6}$$

şeklinde ve sistemin bir üretme olasılığı da yine aynı yaklaşımla,

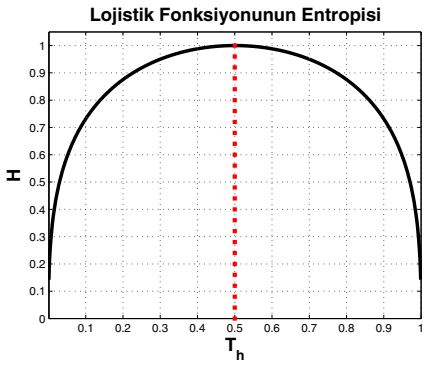
$$\begin{aligned} P(1) = P(x(n) \geq T_h) &= 1 - P(0) \\ &= 1 - \frac{2}{\pi} sin^{-1}\sqrt{T_h} \end{aligned} \tag{7}$$

olarak T_h parametresi cinsinden ifade edilebilir.

Dinamik sistemden en yüksek düzeyde rastsallık elde edebilmek için sistemin ürettiği sayıların entropisinin en yüksek olduğu seviyeye karşı gelen T_h parametresinin değerine göre örnekleme yapılmalıdır. Üretilen bir bitlik sayılar gözönüne alındığında sıfır (6) ve bir (7) üretme olasılıkları, $Shannon$'ın,

$$\begin{aligned} H &= -\sum_{i=0}^{1} P(i) log_2 P(i) \\ &= -(P(0) log_2 P(0) + P(1) log_2 P(1)) \end{aligned} \tag{8}$$

ile verilen entropi tanımına yerleştirilip T_h parametresine göre çizdirilirse Şekil 6'daki entropi - T_h grafiği elde edilir. Grafikte entropinin en yüksek olduğu T_h değerinin 0.5 olduğu gözlemlenebilir. Lojistik fonksiyonuna dayanan GRSÜ modeli $T_h = 0.5$ için eş olasılıklı $(P(0) = P(1))$ sayılar üretebilir.



Şekil 6: Lojistik fonksiyonu için entropi - T_h grafiği.

İdeal bir GRSÜ'nün ürettiği ardışıl iki sayı arasında istatistiksel bağımsızlık bulunmalıdır. Lineer olmayan dinamik sistemlerde ideal koşullar altında üretilen örnekler arasında lineer olmayan belirlenimci bir ilişki mevcuttur. Ancak örneklerin eşik değeriyle karşılaştırılması sonucu üretilen sayılar, T_h'in belli bir değeri için istatistiksel açıdan bağımsız olabilirler [6]. T_h ile ikiye bölünen faz portresinde ardışıl iki bit için sıfır ve bir üretme olasılıkları birikimli olasılık dağılım fonksiyonu cinsinden,

$$\begin{aligned} P_{00}(T_h) &= P(x(n) \in [0, T_h), x(n+1) \in [0, T_h)) \\ P_{01}(T_h) &= P(x(n) \in [0, T_h), x(n+1) \in [T_h, 1]) \\ P_{10}(T_h) &= P(x(n) \in [T_h, 1], x(n+1) \in [0, T_h)) \\ P_{11}(T_h) &= P(x(n) \in [T_h, 1], x(n+1) \in [T_h, 1]) \end{aligned} \quad (9)$$

biçiminde ifade edilebilir. Lojistik fonksiyonu incelenirse, fonksiyonun sabit noktası olan 0.75 değerine göre birleşik olasılıkların parçalı olarak ifade edilebileceği görülür. Birleşik olasılık dağılımları (5)'deki marjinal birikimli olasılık dağılım fonksiyonunu cinsinden Tablo 1 ve 2'deki gibi ifade edilebilir. Dikkat edilirse $T_h = 0.5$ değeri için her bir eleman 0.25 değerini almaktadır. Yani birleşik olasılıklar marjinal olasılıkların çarpımına eşit olmakta, ardışıl üretilen iki bitin, istatistiksel olarak birbirinden bağımsız olduğunu göstermektedir. $T_h = 0.5$ değeri için GRSÜ modeli eş olasılıklı ve bağımsız sayılar üretebildiği için lojistik fonksiyonu bir gerçek rastsal sayı üreteci olarak kullanılma potansiyeline sahiptir.

Tablo 1: $T_h < 0.75$ için marjinal ve birleşik olasılık dağılımları

	P_0	P_1	$x(n)$
P_0	$\frac{1}{2}F_x(T_h)$	$\frac{1}{2}F_x(T_h)$	$F_x(T_h)$
P_1	$\frac{1}{2}F_x(T_h)$	$1-\frac{3}{2}F_x(T_h)$	$1-F_x(T_h)$
$x(n+1)$	$F_x(T_h)$	$1-F_x(T_h)$	1

Tablo 2: $T_h \geq 0.75$ için marjinal ve birleşik olasılık dağılımları

	P_0	P_1	$x(n)$
P_0	$2F_x(T_h)-1$	$1-F_x(T_h)$	$F_x(T_h)$
P_1	$1-F_x(T_h)$	0	$1-F_x(T_h)$
$x(n+1)$	$F_x(T_h)$	$1-F_x(T_h)$	1

5. İSTATİSTİKSEL TEST SONUÇLARI

MATLAB ortamında gerçeklenen modelden üretilen ve ikili tabanda dosyaya kaydedilen 100 milyon bitlik ham GRSÜ verisine NIST 800.22 istatistiksel testleri uygulanmıştır. Elde edilen sonuçlar Tablo 3'de verilmiştir. P-değeri kolonundaki değerler test edilen ham verinin ideal bir gerçek rastsal sayı üretecinden üretilmiş olma olasılığını göstermektedir. Orantı kolonu ise ham veri 1 Mbitlik bloklara ayrılarak test edildiğinde blokların testlerden geçme oranını vermektedir. NIST 800.22 tanım dökümanına göre P-değeri > 0.01 için ham veriyi üreten sistem GRSÜ olarak kabul edilmektedir [7].

Tablo 3: NIST İstatistiksel Test Sonuçları

Test	P-değeri	Orantı
Frekans	0.405613	0.9875
Blok Frekans	0.976574	0.9925
Birikimli Toplamlar	0.663130	0.9850
Koşular	0.600729	0.9950
Uzun-Koşu	0.847183	0.9875
Rütbe	0.465415	0.9900
FFT	0.941144	0.9875
Evrensel	0.709558	0.9875
Yaklaşık Entropi	0.388127	0.9825
Seri	0.917870	0.9925
Doğrusal-Karmaşıklık	0.855534	0.9950

6. SONUÇLAR

Tek boyutlu kaotik fonksiyonların rastsal sayı üretimine uygunluğunu incelemek için lojistik fonksiyonu örnek alınarak geliştirilen matematiksel GRSÜ modelinden hangi şartlar altında eş olasılıklı ve bağımsız rastsal sayıların üretilebildiği belirlenmiştir. MATLAB ortamında gerçeklenen modelin ürettiği sayılar, literatürde rastsal sayı üreteçlerinin değerlendirilmesinde standart kabul edilen NIST 800.22 testlerinden başarıyla geçerek geliştirilen modelin geçerliliğini ortaya koymuştur. Yapılan çalışmada geliştirilen çerçeve yöntemin GRSÜ donanım tasarımı sürecinde devre parametrelerinin belirlenmesinde kullanılması öngörülmektedir.

7. KAYNAKÇA

[1] C. Petrie, J. Connelly, "A noise-based IC random number generator for applications in cryptography," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on,* vol. 47, pp. 615–621, 2000

[2] H. Bock, M. Bucci and R. Luzzi, "An Offset-Compensated Oscillator-Based Random Bit Source for Security Applications," *Cryptographic Hardware and Embedded Systems - CHES 2004*, Vol. 3156, pp. 27–83, 2004.

[3] P. F. Verhulst, "Recherches sur la loi d'mathematics accroissement de la population.", *L'Académie Royale de Bruxelles et de l'Université Louvain,* Vol.18, pp.1–42, 1845.

[4] R. M. May, "Simple mathematical models with very complicated dynamics," *N*ature, Vol. 261, No. 5560., pp. 459–467, 1976

[5] A. Lasota, M. C. Mackey, "Chaos, fractals, and noise: Stochastic aspects of dynamics. Second edition." *A*pplied Mathematical Sciences, 97. Springer-Verlag, New York, 1994.

[6] A. J. Lawrance, R. C. Wolff, "Binary time series generated by chaotic logistic maps." *Stochastics and Dynamics,* 3(4), pp. 529-544., 2003

[7] A. Rukhin et al, "A statistical test suite for random and pseudo random number generators for cryptographic applications," *NIST 800-22rev1a,* 2010.