İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

DOĞAL DİLDE STEGANOGRAFİ

YÜKSEK LİSANS TEZİ

Osman Boyacı

Bilgisayar ve Bilişim Mühendisliği Fakültesi

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Ahmet Cüneyd TANTUĞ

Mayıs 2017

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

DOĞAL DİLDE STEGANOGRAFİ

YÜKSEK LİSANS TEZİ

Osman Boyacı
(504141521)

Bilgisayar ve Bilişim Mühendisliği Fakültesi

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Ahmet Cüneyd TANTUĞ

Mayıs 2017

İTÜ, Fen Bilimleri Enstitüsü'nün 504141521 numaralı Yüksek Lisans Öğrencisi Osman Boyacı, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "DOĞAL DİLDE STEGANOGRAFİ" başlıklı tezini aşağıdaki imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı :	**Doç. Dr. Ahmet Cüneyd TANTUĞ** İstanbul Teknik Üniversitesi	.....
Jüri Üyeleri :	**Yrd. Doç. Dr. Yusuf YASLAN** İstanbul Teknik Üniversitesi	.....
	Doç. Dr. Arzucan ÖZGÜR Boğaziçi Üniversitesi	.....
		.....

Teslim Tarihi : **5 Mayıs 2017**
Savunma Tarihi : **5 Haziran 2017**

Eşime ve aileme

ÖNSÖZ

Bu tezin oluşmasında emeği geçen sayın Doç. Dr. Ahmet Cüneyd Tantuğ'a, beni yeni fikirlerle hep destekleyen iş arkadaşlarıma ve her koşulda yanımda olan eşime ve aileme teşekkürü bir borç bilirim.

Mayıs 2017 Osman Boyacı

İÇİNDEKİLER

Sayfa

KISALTMALAR

DS : Dilbilimsel Steganografi
WSD : Word Sense Disambiguation (Anlam Belirsizliği Giderme)
PL1D : Piecewise Linear 1D Map (1. Dereceden Lineer ve Parçalı Kaotik Harita)
PSD : Power Spectral Density (Spektral Güç Yoğunluğu)
LE : Lyapunov exponent (Lyapunov Üsteli)

SEMBOLLER

DS	**:** Dilbilimsel Steganografi
C	**:** Örtü Metni
S	**:** Örtülü Metin

ÇİZELGE LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

DOĞAL DİLDE STEGANOGRAFİ

ÖZET

Bilgi gizleme insanlığın başlangıcından bu yana önemini hiç yitirmemiş bir konudur. Bu önem sayısallaşan dünyada veri aktarım hızlarının ve depolama alanların artmasıyla daha da açığa çıkmaktadır. Sayısız miktarda veri ses, metin, görüntü ve video olarak internet ortamında sürekli dolaşmaktadır ve bu dolaşım mahremiyet ve gizlilik sorunlarını beraberinde getirmektedir.

Kriptografi verileri şifreleyerek belli bir düzeye kadar bu sorunlara cevap vermektedir fakat kriptografide haberleşmenin gerçekleştiği kanalın varlığı aşikardır. Haberleşme kanalını dinleyen kişiler şifreli bilgiye haiz olamasalar bile veri akışı olduğunu farkedip kanalı kolayca kopabilirler veya akan veriyi değiştirebilirler. Gizli veri iletişimin olduğu kanalın varlığını gizlemek için ise steganografiye ihtiyaç duyulur.

Bu tezde dilbilimsel steganografi yöntemlerinden biri olan eşanlamlı sözcük değiştirme yöntemi kullanılarak Türkçe metinler üzerinde gizli veri iletişimini sağlayan bir steganografi uygulaması tasarlanmıştır. Literatüre katkı olarak gizli veriyi taşıyacak olan Türkçe'deki eş ve yakın anlamlı sözcükler incelenmiş ve aynı anlama gelenler kümelenerek bir sözlük oluşturulmuştur. Diğer bir katkıda sistemin güvenliğini artırmak amacıyla tasarlanan kaotik bir kodlayıcıdır. Son kullanıcılardan alınan dönütlerle sistem değerlendirilmiş ve iyileştirmeler yapılmıştır.

NATURAL LANGUAGE STEGANOGRAPHY

SUMMARY

Information hiding is a vital topic for mankind from the beginning which has never lost its importance. This importance is become more crucial in the digital word by the increasing the data communication rates and storage capacity.

Numerous amounts of data in audio, text, image and video formats are circulating on the internet and this circuiaton brings confidentiality and privacy issues.

Cryptograpy provides security by encrypting the data on a a certain level to those problems but in cryptography the existence of the channel that the communication is performed is obvious. The people who listen to the communication channel can easily break off the channel or change the data even if they do not extract the information. To hide the existence of the communication channel Steganography is needed.

In this thesis, by using synonym replacement technique a linguistic steganography application is designed on the text in Turkish. As a contribution to the literature, synonymous and close meaning words studied, analyzed and clustered together and a thesaurus dictionary was formed by classifying those clusters. Another contribution is proposed chaotic encoder designed to increase the security of the system. Evaluation is made by the feedbacks from end users and improvements were made.

1. GİRİŞ

Antik çağlardan beri gizli bilgi iletimi hep revaçta olmuştur. Uygularlıklar günün getirdiği en iyi teknolojileri bu amaçta kullanıp bir adım öne geçmeyi planlamışlardır. Eski Yunanda balmumu ile kaplanıp üzerindeki metnin okunmasını engelleyen tabletler, elmaların kasada dizilme sırasına göre iletilen gizli mesajlar, görünmez mürekkeple yazılmış yazılar, mors kodlarına gömülmüş mesajlar ve mikro noktalarla iletilen haberler bu amaçla hazırlanmış örneklerden sadece birkaçıdır.

Teknolojinin beraberinde getirdiği bilgiye ulaşım kolaylığı ve işlem güçlerinin artması ise mahremiyet ve güvenlik sorunlarını daha da derinleştirmiştir. Sürekli dinlenen kanallarca bireylerin okudukları her metin, dineledikleri her müzik ve izledikleri her video takip edilmeye başlanmış ve gizli bilgi iletimi önceki çağlara göre çok daha can alıcı hale gelmiştir.

1.1 Steganografi

Kökleri antik Yunan'a kadar uzanan staganografinin anlamı örtülü, gizli metindir. Yunanca gizlemek, örtmek manasında gelen $\sigma\tau\varepsilon\gamma\alpha\nu o\zeta\,(steganos)$ ve yazma anlamına gelen $\gamma\rho\alpha\phi\varepsilon\iota\upsilon(graphein)$ sözcüklerinden alır. Steganografininin temel amacı iletilmek istenen bilginin varlığını gizleyip, sadece anahtara sahip olan kişilerce algılanmasını sağlamaktır.

Ses, görüntü ve video gibi alanlarda steganografik teknikler sinyal işleme teknikleriyle paralel bir gelişim göstermiş ve kullanımı yaygınlaşmıştır. Resim içine resim gömme, ses sinyalindeki gürültüye gizli veri yükleme, video kareleri arasına fazladan bir kare ekleme gibi basit teknikler bile insan algısının farkedebilme eşiğini geçmediğinden bu alanlarda steanografi hızlı bir gelişim göstermiştir. Diğer yandan sinyal işleme teknikleri metin üzerinde kullanılmaya çalışıldığında değişiklik insan gözü tarafından kolayca farkedilir çünkü metinler dildeki belli bir gramer kuralınca belli sentaktik ve semantik kurallarını yerine getirerek anlamlı bir bütün oluşturur. Oluşabilecek herhangi bir aykırılık steganografik uygulamayı başarısızlığa uğratacaktır. Bütün bu

sebeplerden dolayı dilbilimsel steganografi diğer steganografi dallarına göre daha az ve yavaş bir gelişim göstermiştir ve bu alanda alınacak daha çok yol vardır.

1.1.1 Dilbilimsel Steganografi

Dilbilimsel steganografi doğal dildeki bir metin üzerinde yapılan değişikliklerle karşı tarafa bilgi göndermeyi amaçlayan steganografinin alt dallarından biridir. Diğer steganografik alanların aksine üzerinde çok çalışılmamış olmasına rağmen doğal dil işleme tekniklerinin gelişmesiyle literatürde kendine yer bulmaya başlamıştır.

Ses, görüntü ve video gibi veriler çağımızda popülerliğini artırmasına rağmen hayatımızda ondan çok daha eski ve önemini hiç yitirmeyecek bir ortam daha vardır: metinler. Doğal dilde yazılmış milyonlarca metin bilgi gömme aracı olarak mükemmel bir ortam sunmaktadır ve bu tezde bunlardan yararlanılmıştır. Farklı dilbilimsel teknikler incelenmiş ve Türkçe üzerinde daha önce çalışılmamış bir teknik olan eş anlamlı kelimelerin bilgi taşıyıcısı olarak kullanıldığı bir stegosistem tasarlanmıştır.

1.1.2 Dilbilimsel Stegosistem

Doğal dilde oluşturulmuş bir metindeki cümlelerin sentaktik, semantik ve sözcüksel yapısında değiştirmelerek yaparak gizli haberleşmeyi sağlayan sistemlere dilbilimsel stegosistemler adı verilir.

Hangi yöntemi kullanırsa kullansın, bir dilbilimselin amacı sentaktik, semantik ve sözcüksel bazda doğal dile mümkün olduğunca yakın; stego anahtara sahip kişilerce gizli bilginin çıkarılmasına imkan verdiği halde anahtara sahip olmayan kişiler tarafından hiç bir şüphe çekmeyen metinler üreterek gizli iletişimi sağlamaktır.

Bu amaçla bu çalışma kapsamında Türkçe için mevcut yöntemler incelenmiş, mevcut çözümlerin artıları ve eksileri detaylı araştırılarak yeni bir çözüm önerilmiştir.

1.2 Literatür Araştırması

Steganografinin geçmişi antik Yunan'a kadar gitmesine rağmen ilk bilimsel çalışma Simmons'un 1984'te ortaya attığı ünlü mahkum problemiyle başlar [1] [2]. Bu problemde hapishanede birbirlerinden ayrı odalarda tutulan iki mahkum arasında gardiyanın aracılığıyla iletilen mesajlarla bir kaçış yapılıp yapılamacağı üzerinedir.

Yapılan çalışma örtülü kanal yaratmanının varlığından bahsedilen ilk çalışmalar olması açısından oldukça önemlidir. Daha sonra Bağlamdan Bağımsız Gramer (BBG) tekniğini ve Huffman kodunu kullanan Wayner ilk çalışmasında "mimik fonksiyonarıyla" bir metindeki istatistiği başka bir metinde taklit edip gizli bir kanal oluşturmaya çalışır [3]. İkinci çalışmasında ise yine aynı metodlarla pratik bir uygulama ortaya koyar ve bahsi geçen mimik fonksiyonlarının RSA kadar güvenlik sağladığını açıklar [4]. Stegonagrafi alalında uluslarası bir çalıştay düzenlendikten sonra steganografinin ne olduğu ve neler yapabileceği, ortak terimlerin kullanımı ve genel teorisi ortaya konulur ve bu alanda yapılan çalışmalar hız kazanır [5].

Chapman tezinde şifrelenmiş veriyi şüphe çekmeyecek biçimde gizlemek için özel olarak hazırlanmış sözlüklerle BBG kullanır [6]. Daha sonra bu çalışmalarını içeriğe göre hazırlanmış şablonlar ve eşanlamlı sözcükleri kullanacak şekilde genişletir ve büyük metinlerde de gizli veri saklayacabilecek hale getirir [7].

Cachin 1998'de yayınladığı makalesinde konuyu bilgi teorisi kapsamında ele alır ve kanalı dinleyen üçüncü kişilerce gizli bilgi alışverişinin sezilmesini hipotez testiyle açıklar [8]. Güvenli stego sistemlerin varlığını ortaya koyar ve örtülü metindeki dağılımın belli şartları sağladıktan sonra göreli entropiyle bu sistemlerin güvenli olduğunu kanıtlar. Ortaya koyduğu model ve konuyu bilgi teorisi penceresinden ele alıp literatüre yaptığı katkıların önemi çok büyüktür.

Steganografinin teorisi ortaya atıldıktan ve güvenli stegosistemlerin varlığı ispatlandıktan sonra Petitcolas bu alanda uygulanan tekniklerin eskiden beri var olduğunu, neyin işe yarayıp yaramayacağını ve hangi konuların üzerinde durulması gerektiğini özetleyen bir çalışma yayınladı [9]. Bu çalışmasında mevcut yöntemleri geçmiş yöntemlerle birlikte ele alıp sınıflandırdı ve dikkat edilmesi greken noktaları madde madde işleyip kendisinin de önerdiği saldırılarla sistemdeki zayıflıkların nasıl ortaya çıkarılabileceğini ortaya koydu. Yapılan çalışmaları özetlemesi ve geçmişteki sistemlerden ders alınması gerektiğini ortaya koyması bakımından bu alandaki önemli bir makaledir.

Purdue CERIAS'ta Atallah ve ekibi ağırlıklı olarak metin damgalama olmak üzerinde bu alanda literatüre bir çok katkıda bulunmuştur. [10] ve [11]'de, doğal dilde damgalama için bir şema önerilmiş, [12]'de ise steganografi ve steganaliz için

gürbüzlük ve karşılan sorunlar ele alınarak özet bir çalışma yayınlanmıştır. Bu çalışmasında Bennett ileride anlamsal bazda yapılacak değişikliklerle daha fazla yol katedilebilineceğini ortaya koyar. [13]'da ise çeviri bazlı bir tasarımdan bahsedilir. Bir metni farklı bir dile çevirme işleminde birden fazla seçenekle karşılaşılır ve bu seçenekler kodlanarak bilgi gömülmeye çalışılır. [14]'de ise "Niceliksel Esneklik Ölçütü" kavramıyla örtü metninde meydana gelen bozulmalar ölçülür ve eşanlamlı kelime seçiminde hep en anlamı en belirsiz olan seçilir oluşacak olan örtülü metin maksimum belirsizlik içersin ve mümkün olduğunca bozulma limitine yakın olsun. Hatta öyleki üçüncü kişilerin metindeki damgaya zarar vermesini önlemek amacıyla damgalanmak istenen bitler bitse bile veri gömmeye devam edilir. Grubun bir sonraki makalesinde ise damgalama için sadece kelimelerin kullanılmasının yeterli gelmediği ifade edilir ve cümle bazında damgalama yapılır [15]. Topkara'nın bir diğer çalışmasında ise doğal dilde damgalama yaparken karşılaşılan zorluklar anlatılır [16]. Damgalamanın kalitesini ölçmek üzere istatistiksel makine çevirisinde sıklıkla kullanılan BLEU metriği kullanılır. Bu metriğe göre makalade tasarlanan sistemin başarısı 1 üzerinden 0.45 BLEU olarak ölçülmüştür. Çeviri tabanlı diğer bir çalışmada ise damgalama orijinal metne ihtiyaç duymadan, sadece damgalı metinle damga çıkaralabilmesi açısından çeviri bazlı çalışmalar bir adım öteye götürülmüştür [17]. Grubun bir sonraki makalesi ise staganaliz üzerinedir. Sözcük tabanlı örtülü metinlerde evrensel steganaliz ve "Destek Vektör Makineleri" yöntemleriyle bilgi gömülü verileri diğerlerinden ayırmadaki başarım %84.9 olarak ölçülmüştür [18].

Bolshakov çalışmasında özel olarak hazırlanmış eşanlamlı bir sözlük ve kelimelerin eşdizimsel istatistiğini tutan bir veri tabanıyla Ruşça özelinde bir stagosistem tasarlamıştır [19]. Bu sözlükte mutlak eş anlamlılar bağlamdan bağımsız olarak kullanılırken göreli eş anlamlılar bağlama bakılarak kullanılmıştır.

Bergmair [20]'de dilbilimsel steganografiyi sistematik olanlar inceleyip bu alandaki yaklaşımları, sistemleri ve ve sorunları özetlemiş, [21]'de ise bu alanda yapılmış çalışmaları derlemiştir. Yazar diğer bir çalışmasında ise dilbilimsek steganografiyi HIP (Human Interactive Proof) yaklaşmıyla ve WSD'yi kullarak Turing testini tersine kullanmayı önermiştir [22].

Türkçe özelinde çalışma yapan tek grup ise Sankur ve ekibidir. Yayınladıkları makalelerde Türkçe için cümle bazında 20 farklı sentantik değiştirme önerilmiş ve

bu değiştirmelerle damgalama yapılmıştırç [23] [24] [25]. Çalışmadaki veri gömme oranı cümle başına 0.5 bit ile 1 bit arasında değişirken bu oran güvenlik gerektiren uygulamalarda 0.1'in altına düşmektedir.

1.3 Motivasyon ve Katkılar

Literatürdeki mevcut çözümler ve sistemler incelendikten sonra bu çalışmalaya aşağıdaki motivasyonlarla yön verilmiştir:

- **Literatürdeki dilbilimsel steganografi çalışmaların azlığı**
  Steganografik alanda yapılan çalışmaların büyük çoğunluğu görüntü ortamı içindir. Genellikle sinyal işleme teknikleri kullanılarak yapılan bu çalışmaların dilbilimsel steganografiye katkısı çok azdır.

- **Pratik uygulamaların azlığı**
  Yapılan çalışmalar genellikle kağıt üzerinde kalmakta ve pratik bir değeri olmamaktadır. Labaratuvar ortamında belli özel koşullarda çalışan tasarımlar pratikte kullanım bulmamaktadır.

- **Mevcut yöntemlerdeki güvenlik açıkları**
  Her ne kadar üzerinde çalışılan veriler doğal dilde ve kullanılan teknikler doğal dil işleme teknikleri içinden olsa da nihai sistemler bir güvenlik sistemidir ve güvenliği sağlamak amacıyla tasarlanmıştır. Bu bakımdan önerilen sistemlerde güvenlik baştan aşağı hiç göz ardı edilmemesi gereken bir konudur. Untulmamalıdır ki bileşenlerden herhangi birindeki bir güvenlik açığı bütün sistemi zayıf hale getirebilir.

- **Türkçe için mevcut çalışmaların azlığı**
  Türkçe için DS alanında yapılmış sadece tek bir çalışma göze çarpmaktadır. Bu çalışmada daha önce bölüm1.2'de bahsedildiği gibi metin üzerinde cümle bazında sentaktik değişiklikler yapılarak metin damgalama yapılmaktadır. Diğer tekniklerle ilgili literatürde hatrı sayılır bir çalışma yer almamaktadır.

- **Türkçe'nin eklemli ve fazla özellikli yapısı**
  Türkçe sonradan eklemeli yapısı gereği bir çok özellik barındırır. Bu da steganografi tasarımcısına bilgi gömebileceği geniş bir alan bırakmaktadır.

- **Türkçe'de sistematik bir eşanlamlı sözlük verisinin olmayışı**
  Sözlük tabanlı stegosistemlerin olmazsa olmazı eş ve yakın anlamlı sözlüklerdir. Wordnet benzeri anlam haritasının çıkarıldığı bir sözlük var olmadan güvenli bir sistem tasarlamak zordur çünkü yapılan değişiklikler insan gözüne çarpacaktır.

Tez kapsamında bahsedilen eksik noktalar tespit edilip dilbilimsel stegosistem tasarlanmış ve bunların literatüre katkılarının aşağıdakiler olması beklenmektedir:

- **Pratik bir dilbilimsel steganografi örneği**
  Son kullancıların da faydalanabileceği kullanımı basit pratik bir DS uygulaması tasarlanmıştır.
- **Yeni bir kodlayıcı**
  Kaotik dinamik sistemleri kullanan yeni bir rastgele sayı üreteci ve kodlayıcısı tasarlanmış ve steganografik sistemlerdeki kullanımı anlatılmıştır.
- **Türkçe için steganografik bir sistem**
  Türkçe'nin steganografik uygulamalarda da kullanılabilineceği yapılan tasarımla ortaya konulmuştur.
- **Türkçe eş ve yakın anlamlılar sözlüğü**
  Doğal dil işlemede steganografi, steganaliz yeniden yorumlamada(paraphrasing) kullanılmak üzere sistematik bir eş ve yakın anlam sözlüğü derlenmiştir.

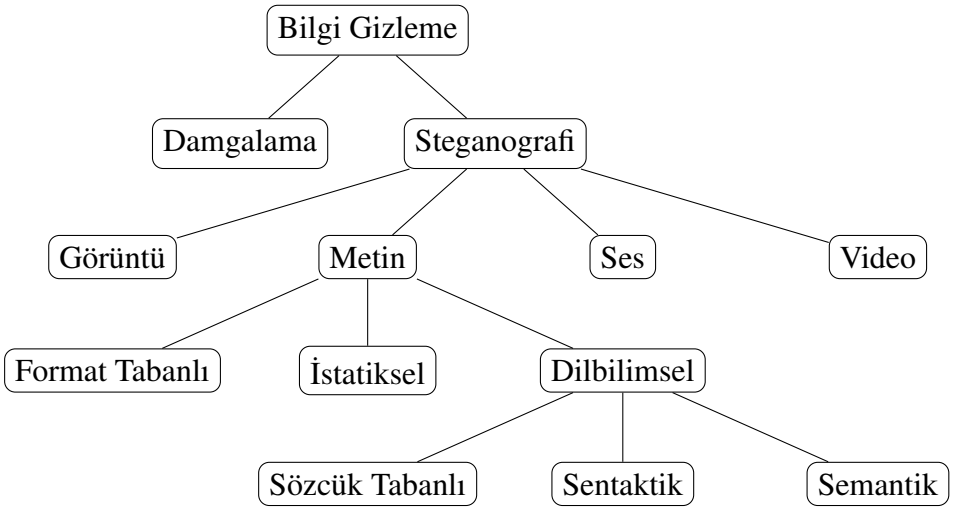
1.4 Genel Bakış ve Organizasyon

Tezin 2. bölümünde bilgi gizleme tanımları ve ayrıntılarıyla verilecek, steganografinin kriptografiden ve damgalamadan farkları anlatılıp farklı yöntemlerden ve bunları değerlendirmek için gereken metriklerden bahsedilecektir. 3. ve 4. bölüm önerilen sisteme ayrılmıştır. Kullanılan doğal dil işleme teknikleri, sözlükten eş anlamlı kelime arama işlemleri, bulunan eş anlamlı sözcüklerin gizlenmek istenen veriyi temsil edecek şekilde kodlanması, kodlanan sözcüklerin metne nasıl gömüleceği ve nasıl çıkartılacağı bu bölümde anlatılmıştır. Tasarlanan sistemin performansının değerlenmdirilmesi 5. bölümde ele alınacaktır. Gerçek kullanıcılardan alınan dönütler ve sistemdeki iyileştirmelerden bahsedilecektir. Son olarak 6. bölümde ise tasarım genel olarak değerlendirilip sonuçlar tartışılacak ve önerilere yer verilecektir.

2. BİLGİ GİZLEME

2.1 Tanımlar ve Genel Çerçeve

Bilgi gizleme bilimi çağlar boyunca gelişerek Şekil 2.1'de verilen alt bileşenlere sahip olacak şekilde evrimleşmiştir.

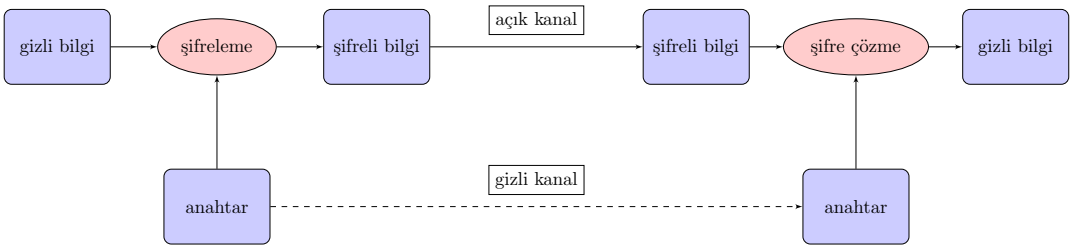


Şekil 2.1 : Bilgi Gizleme ve alt sınıfları.

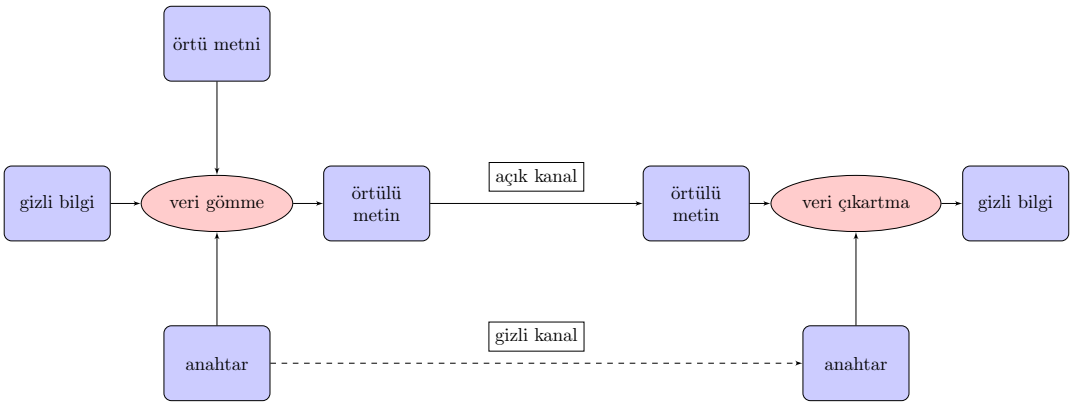
Bilgi gizlemede kriptogtafi ve steganografi çok karıştırılan bir konudur. En genel anlamda kriptografinin temel amacı açık veriyi şifreli hale getirip üçüncü kişilerin açık veriye haiz olmasını engellemektir. Şifreli verinin ortada durması kriptografi açısından bir sorun teşkil etmez. Kriptolojinin aksine, stganografi bu verinin varlığını gizleyip iletişimi üçüncü kişilerden habersiz hale getirmeye çalışır. Şekil 2.2'de genel bir kripto uygulamasının ana şeması, Şekil 2.3'de ise genel bir kripto uygulamasının ana şeması verilmiştir. Bu iki şema arasında en temel fark steganografinin bilgiyi gizleyecek bir örtü metnine ihtiyaç duymasıdır. Çizelge 2.1'de bu iki teknik arasındaki fark ayrıntılı olarak verilmiştir [26].

Karıştırılan diğer bir konuda damgalama ve steganografidir. Çizelge 2.2'de bu iki konuya açıklık getirilmiştir [26].

Hangi alanda yapılırsa yapılsın, hangi teknik kullanılırsa kullanılsın bir stegonagrafi uygulamasında temel bileşenler şunlardır [27]:



Şekil 2.2 : Genel bir kriptografi uygulamasının şeması.



Şekil 2.3 : Genel bir stegonagrafi uygulamasının şeması.

Örtü Nesnesi (Cover Object): İçinde henüz gizli bilgi barındırmayan, bilginin gömüleceği nesnedir. Örnek: bir video karesi.

Bilgi taşıyıcı (Informatin Carrier): Örtü nesnesinde var olan, bilginin asıl yükleneceği alt birimlerdir. Örnek: resimdeki bir piksel.

Gizli Bilgi (Secret Information): Örtü nesnesindeki bili taşıyıcılarınca karşı tarafa iletilmek istenen bilgidir. Örnek: "Şafak operasyonu başlasın".

Örtme Anahtarı (Stegokey): Örtü nesnesine veri gömülürken ve veri çıkartılırken kullanılan gizli anahtardır. Sadece bu anahtara sahip olan kişilerce gizli veri anlamlı hale gelir. Örnek : "0xDEADBEEF".

Kodlama (Encode): Gizli bilginin örtme anahtarı yardımıyla örtülü nesnedeki bilgi taşıyıcılarınca nasıl temsil edileceğine karar verme işlemidir. Örnek: gizli bilgi 1 ise kodlanmış bilgi 0, 0 ise kodlanmış bilgi 1.

	Kriptografi	**Steganografi**
Amaç	İçeriği karartmaktır.	İletişimi gizlemektir.
Gizlilik	Şifreli veri anlamsızdır.	Gizli veri *görünmezdir.*
İetişimin güvenliği	Anahtarın gizliliğidir.	Veri gömme metodudur.
Sağlamlık garantisi	Şifreleme algoritmasıdır.	İstatiksel ve algısal görünmezliktir.
Saldırılar	Bulmak kolay, çıkartmak zordur.	Bulmak zor, çıkartmak zordur.
Önlemler	Tersine mühendislik yapmaktır.	Verileri sürekli dinleyip istatistik yapmaktır.

Çizelge 2.1 : Kriptografi ve Steganografi.

	Damgalama	**Steganografi**
Amaç	Damgayı korumaktır.	Gizli bilgiyi ifşa etmemektir.
Gizlilik	Şifreli veri anlamsızdır.	Gizli veri *görünmezdir.*
Sağlamlık	Kurcalamaya ve kaldırmaya karşıdır.	Ortaya çıkarmaya karşıdır.
Sinyal işleme, rastgele hatalar, sıkıştırma	Damgayı yok etmemelidir.	Gizli veriyi kaybetmeye yol açabilir.
Taşıyıcı türü	Sayısal verilerdir (görüntü, video, ses, metin).	Herhangi bir servis, protokol veya dosyadır.

Çizelge 2.2 : Damgalama ve Steganografi.

Veri Gömme (Embedding): Kodlanmış gizli bilgiyi örtme anahtarı yardımıyla örtü nesnesindeki bilgi taşıyıcılarına yükleme işidir. Örnek: resimdeki bir pikselin son bitine kodlanmış veriyi yüklemek.

Örtülü Kanal (Cover Channel): Gizli iletişimin kurulacağı, üçüncü kişilerce dikkat çekmeyen ortamdır. Örnek: bir televizyon kanalı.

Örtülü Nesne (Stego Object): İçinde gizli bilgiyi barındıran, bilginin gömüldüğü nesnedir. Örnek: pikselleri değiştirilmiş bir resim

Veri (Çıkarma Extracting): Kolanmış gizli bilgiyi örtme anahtarı yardımıyla örtülü nesnedeki bilgi taşıyıcılarından çıkarma işidir. Örnek: resimdeki bir pixelin son bitlerini okumak.

Kod çözme (Decode): Örtülü nesnedeki bilgi taşıyıcılarından çıkartılan kodlanmış verinin çözülme işlemidir. Örneğin gizli bilgi 1 ise kodlanmış bilgi 0, 0 ise kodlanmış bilgi 1.

Aktif bekçi (Aktive Warden): Örtülü kanalı dinleyen ve iletişimdeki verileri değiştiren üçüncü kişi ve kişilerdir. Örnek: resimdeki pikselleri inceleyip değiştiren bir kişi aktif bekçidir.

Pasif bekçi (Passive Warden): Örtülü kanalı dinleyen be iletişimdeki verileri değiştirmeyen üçüncü kişi ve kişilerdir. Örnek: resimdeki pikselleri inceleyen bir kişi pasif bekçidir.

2.2 Kullanılan Yöntemler ve Ortamlar

Steganografide kullanılan yöntemleri en başta verinin gömüleceği ortam belirler. Bu bakış açısıyla steganografi ses, görüntü ve metin olarak farklı kategoride incelenebilir. Bu tez kapsamında Şekil 2.1'de verilen sınıflandırmadan metin bazlı bir uygulama gerçekleştirelceği için yöntemler bu konu özelinde incelenmiştir.

Orjinal cümle	Ayşe tatile␣çıksın.
Format tabanlı	Ayşe tatile␣␣çıksın.
Sözcük tabanlı	Ayşe tatile gitsin.
Sentaktik	Ayşe çıksın tatile.
Semantik	Ayşe biraz dinlensin.

Çizelge 2.3 : Metin tabanlı steganografi örnekleri.

Kullanılan yöntemleri özetlemesi bakımından Çizelge 2.3'de "Ayşe tatile çıksın" cümlesi üzerinde 4 farklı örnek verilmiştir. Alt başlıklar halinde kısaca bunlara değinilecek ve bölüm 3'de tasarlanan stegosistem ayrıntılarıyla ortaya konulacaktır.

2.2.1 Format Tabanlı

Format tabanlı yöntemler genelde metin üzerinde insan gözünün farkedemeyeceği değişikliklerle bilgi gizlenmek istenir. Fazladan boşluk bırakmak, yazı fontunu, büyüklüğünü sayfa boyutunu ve girintilemeyi değiştirmek metin formatlı

steganografinin temel örneklerindendir. Günümüz bilgisayarlarının işlem güçlerinin artmasıyla neredeyse hiç bir güvenlik sunmazlar [27].

2.2.2 Sözcük Tabanlı

Sözcük tabanlı yöntemlerde bilgi taşıyıcıları sözcüklerdir. Gizlenmek istenen bilgi kelimelerde yapılacak değişikliklerle cümlenin sentaktik ve semantik yapısını bozmadan örtü metnine veri gömmeye çalışır. Bu konuda yapılan en önemli çalışmalar eş anlamlı kelime kullanımıdır. Tasarlanan pratik dilbilimsel stegosistemler arasında en yaygın ve iyi sonuç veren yöntem anlamdaş kelimerle bilgi gizlenen yöntemdir çünkü [12]:

(i) **Sentantik yapı bozulmaz.**
Yeni seçilen sözcüklerle orjinalleri aynı tipten olacağından söz dizilimini etkilenmez.

(ii) **Semantik yapı bozulmaz.**
Yeni seçilen sözcükler de aynı anlamda olduğu için cümlede anlamsal herhangi bir değişiklik gözlenmez.

(iii) **İstatiksel yapı bozulmaz.**
Yapılan değişiklik sadece belli şartları sağlayan eş anlamlı sözcüklerde olduğu için ve diğer sözcükler herhangi bir değişikliğe uğramadığı için kullanım sıklıkları neredeyse aynıdır.

Bu sebeplerler bu tez anlamdaş kelimere odaklanır.

2.2.3 Sentantik

Sentantik dilbilimsel steganografi ise cümledeki kelimelerin diziliminde yapılan değişikliklerle veri gizlemeye çalışır. [25]'de verilen çalışmada Türkçe cümeleler için 20 farklı değişiklik önerilmiştir. Cümlelerdeki sıklıkları göz önüne alındığında en önemli örnekleri aktiflik-pasiflik $(\%35)$, zarfın yerini değiştirme $(\%14.8)$ ve bağlanan kelimelerin yerini $(\%12.4)$ değiştirmektir.

Türkçe gibi sözdiziminin görece rahat olduğu dillerde bu teknik daha rahat uygulanabilmesine rağmen güvenlik gerektiren uygulamalarda veri taşıma kapasitesi bakımından yetersiz kalmaktadır (0.1 bit/cümle) [25].

2.2.4 Semantik

Semantik dilbilimsel steganografide ise metnin anlamı aynı olacak şekilde cümleler farklı sözcüklerle ve farklı dizilimde olabilir. Çizelge 2.3'de verilen örnekte anlamsal olarak tatile çıkmakla dinlenmek arasında herhangi bir fark yoktur ve bu alandaki stegosistemler bu tarz yapıları kullanmak isterler.

Ontolojik veriler bu teknikte çok önemli bir yere sahiptir zira cümlede anlamsal herhangi bir değişiklik yapılmak istenildiğinde ilk başvurulacak veriler ontolojilerdir. Son yıllarda yapılan çalışmalar bu konuya odaklansa da teknik henüz olgunlaşmadığından pratik manada kullanımına pek rastlanmaz [28].

2.3 Kullanılan Metrikler

Bir steganosistemde şu metrikler önemlidir: kapasite, dayanıklılık, güvenlik ve veri gömme oranı. Aşağıda alt başlıklar halinde bu metrikler incelenecektir.

2.3.1 Kapasite

Örtü metninin sahip olduğu bilgi taşıyıcılarının yoğunluğudur. Cümle başına veya kelime başına gömülebilecek bit oranıyla ifade edilir.

2.3.2 Dayanıklılık

Örtülü metnin ifşa olmaya karşı ne kadar dirayetli olduğunun bir ölçüsüsüdür. Aktif bekçinin sistemin bir parçaşı olduğu damgalama uygulamalarında daha büyük bir öneme sahiptir.

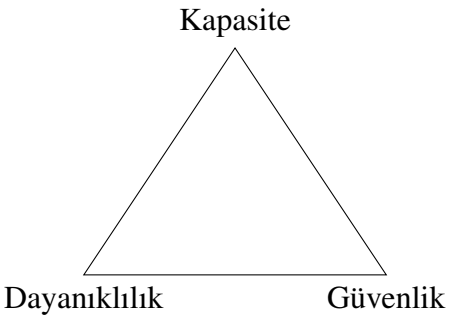
2.3.3 Güvenlik

Gizli bilginin örtülü veride ne kadar *görünmez* olduğunun bir ölçüsüdür. Göreli entropi kavramıyla örtü metnindeki bozulmayla ifade edilir.

2.3.4 Veri Gömme Oranı

Kapasitenin kullanılma oranıdır. Örtü metninin sunduğu bilgi taşıyıcısı başına stegosistemde uygulanan değişkliklerdir. Genellikle maksimum değeri olan 1'e yakın kullanılır.

Bütün mühendislik uygulamalarında olduğu gibi steganografide de da bu metrikler arasında bir ödünleşim (trade-off) vardır. Şekil 2.4 'de bu ilişki gösterilmiştir.



Şekil 2.4 : Steganografik metrikler arasındaki ilişki.

Yine de tüm yönleriyle iyi analiz edilmiş sistemlerle güvenli, dayanıklı ve yüksek kapasiteli bir stegosistem tasarlamak mümkündür. Bir sonraki bölümde bu metrikleri mümkün olduğunca yüksek bir stegosistem tasarımına yer verilecektir.

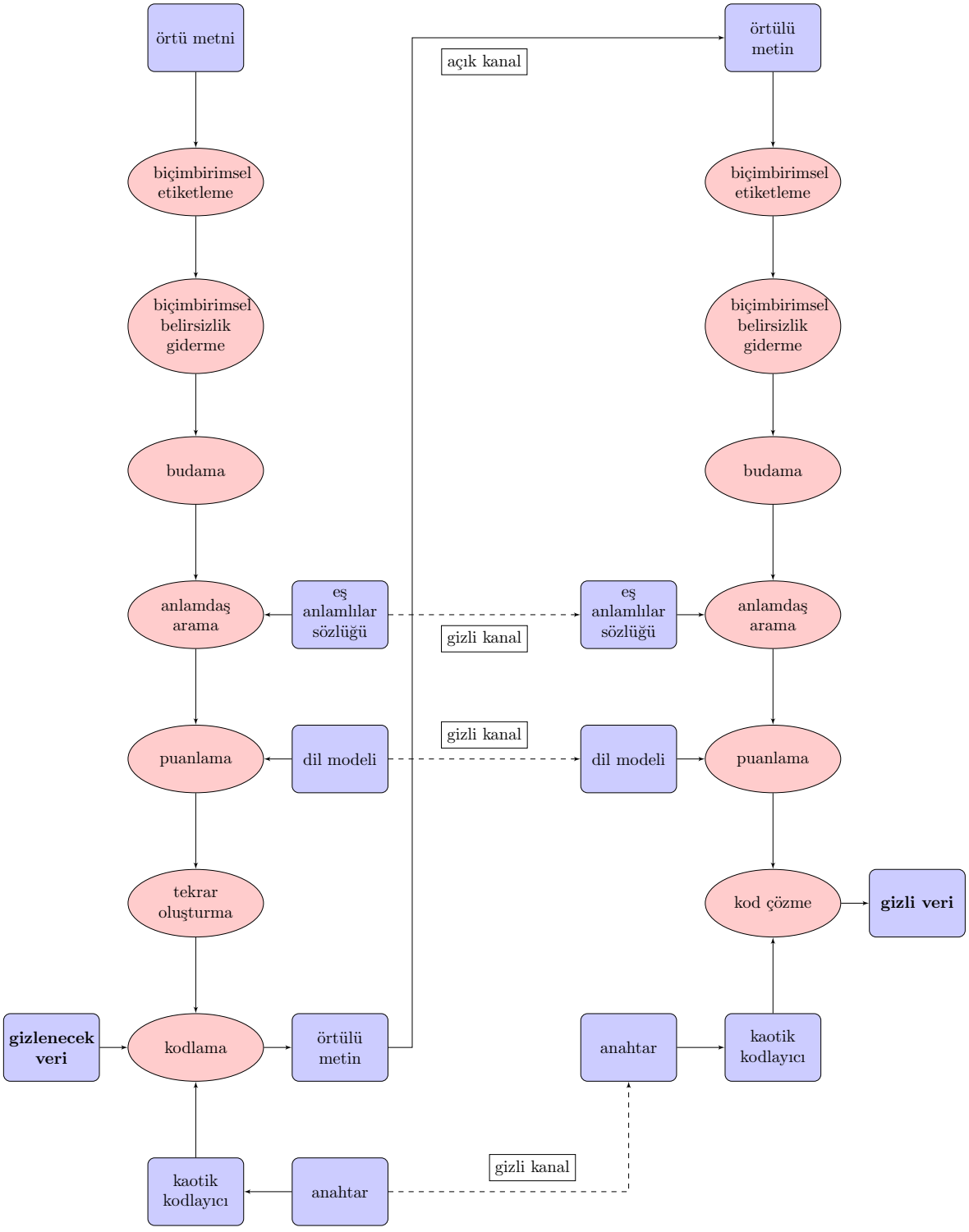
3. ÖNERİLEN DİLBİLİMSEL STEGOSİSTEM

Literatürdeki mevcut çalışmalar araştırılıp varolan teknikler üzerinde yapılan incelemeler sonucu Bölüm 1.3 ve 2.2.2'de açıklanan sebeplerle tez kapsamında Türkçe dilinde sözcük tabanlı bir stegosistem tasarlanmasına karar verilmiştir. Üçüncü kişilerce herhangi bir dikkati üzerine çekmeden, örtü verisinde münkün olan minimum değişikliği yaparak gönderilmek istenen veriyi örtülü veriye gömmek ve gizli bir iletişim kanalı kurmak tasarlanacak stegosistemin başlıca amacıdır. Bu amaç doğrultusunda bilgi taşıyıcı olarak eşanlamlı kelimeler kullanılmış; semantik, sentaktik ve istatistiksel olarak doğal dilde oluşturulan orjinal metinlere mümkün olduğunca yakın örtülü veriler oluşturulması hedeflenmiştir.

Türkçe sonradan eklemeli bir dil olduğundan kelime kökleri yapım ekleri ve çekim ekleri alarak gövdeleri oluşturur. Bu yüzden örtü metninde yer alan bir kelimenin eş anlamlısıyla değiştirmek için önce onu çekim eklerinden arındırmak, sonra sözlükten anlamdaş kelimeyi bulup cümledeki formuna geri getirmek ve uygunluğuna bakmak gerekir. Şekil 3.1'de sistemin genel akış şeması verilmiştir. Alt başlıklar halinde sistemin bütün bileşenleri incelenecektir.

3.1 Biçimbirimsel Etiketleme

Stegosistemde yapılması gereken ilk temel iş doğal dilde oluşturulmuş örtü metnindeki cümleleri alıp kelimelerine ayırmak ve etiketlemektir. Her bir kelime anlamlı en küçük parçasına kadar ayrılmalı ve hangi ekin ne için kullanıldığına karar verilmelidir. Biçimbirimsel etiketleme için Oflazer'in tasarladığı sistem kullanılmıştır [29]. Sonlu durum makinaları kullanılarak tasarlanan bu sistemde kullanılan etiket uzayı Çizelge 3.1'de verilmiştir. Bu etiketleyicide dilde geçen bütün kelimeler kelime köküne bahsedilen uzaydaki etiketler eklenerek çözümleme yapılır. Çizelge 3.2'de "fikirlerin tartışılmadığı yerde şiddet devreye girer" örnek cümlesi için etiketlenme verilmiştir.



Şekil 3.1 : Tasarlanan steganografi uygulamasının şeması.

Sınıf	**Etiket**
Main	"Adj", "Adverb", "Conj", "Det", "Dup", "Interj", "Noun", "Num", "Postp", "Pron", "Punc", "Ques", "Verb"
Adj	"PastPart", "FutPart", "PresPart", "Related", "FitFor", "JustLike"
Noun	"Inf1", "Inf2", "Inf3", "PastPart", "FutPart", "Prop", "Zero"
Num	"Card", "Ord", "Percent", "Range", "Real", "Ratio", "Dist"
Postp	"PCAbl", "PCAcc", "PCDat", "PCGen", "PCIns", "PCNom"
Pron	"Demons", "DemonsP", "Ques", "QuesP", "Reflex", "ReflexP", "Pers", "PersP", "Quant"
Noun2Adj	"With", "Without", "InBetween", "Rel"
Verb2Adj	"Continue"
Adj2Adverb	"Ly"
Noun2Adverb	"Since"
Verb2Adverb	"AfterDoingSo", "SinceDoingSo", "When", "ByDoingSo", "While", "AsIf", "WithoutHavingDoneSo", "WithoutBeingAbleToHaveDoneSo", "Adamantly"
Adj2Noun	"Ness"
Noun2Noun	"Agt", "Dim"
Verb2Noun	"ActOf", "NotAbleState", "NotState", "FeelLike"
Adj2Verb	"Become", "Acquire"
Noun2Verb	"Become", "Acquire"
A123	"A1pl", "A1sg", "A2pl", "A2sg", "A3pl", "A3sg"
P123	"P1pl", "P1sg", "P2pl", "P2sg", "P3pl", "P3sg", "Pnon"
Case	"Nom", "Acc", "Dat", "Loc", "Abl", "Gen", "Ins", "Equ"
Voice	"Pass", "Caus", "Reflex", "Recip"
Modal	"Able", "Repeat", "Hastily", "EverSince", "Almost", "Stay", "Start"
Polarity	"Pos", "Neg"
Tense	"Past", "Narr", "Fut", "Aor", "Pres", "Desr", "Cond", "Neces", "Opt", "Imp", "Prog1", "Prog2"
Verb Optional	"Cop"

Çizelge 3.1 : Etiketlemede kullanılan tüm uzay.

Kelime	Etiket
fikirlerin	fikir $+Noun+A3pl+Pnon+Gen$
	fikir $+Noun+A3pl+P2sg+Nom$
tartışılmadığı	tart $+Verb+Recip\hat{}DB+Verb+Pass+Neg\hat{}DB+Adj+PastPart+P3sg$
	tart $+Verb+Recip\hat{}DB+Verb+Pass+Neg\hat{}DB+Noun+PastPart+A3sg+P3sg+Nom$
	tartış $+Verb\hat{}DB+Verb+Pass+Neg\hat{}DB+Adj+PastPart+P3sg$
	tartış $+Verb\hat{}DB+Verb+Pass+Neg\hat{}DB+Noun+PastPart+A3sg+P3sg+Nom$
yerde	yer $+Noun+A3pl+P2sg+Nom$
şiddet	şiddet $+Noun+A3sg+Pnon+Nom$
devreye	devre $+Noun+A3sg+Pnon+Dat$
girer	gir $+Verb+Pos+Aor+A3sg$
	gir $+Verb+Pos+Aor\hat{}DB+Adj+Zero$
.	. $+Punc$

Çizelge 3.2 : Kelime etiketlemeden örnekler.

3.2 Biçimbirimsel Belirsizlik Giderme

Türkçe metinlerde Çizelge 3.2'de de görüldüğü gibi bir kelimenin birden fazla çözümlemesi olabilir. Bu çözümlemelerden kelimenin o cümlede hangi haliyle geçtiğini anlamak için yapılan işleme Biçimbirimsel Belirsizlik Giderme (Morphological Disambiguation) denir. Tasarımda Yuret'in tasarladığı belirsizlik giderici kullanılmıştır [30]. Bu tasarım cümlede geçen birden fazla etiketli her kelime için cümledeki diğer kelimelerden yararlanarak bir kestirim yaparak doğru çözümlemeyi bulmaya çalışır. Başarım oranı %96'lara kadar ulaşır. Çizelge 3.3'de örnek olarak etiketlenmiş kelimeler ve bunların cümle içindeki cümle içindeki doğru etiketlenmesine yer verilmiştir.

3.3 Budama

Kullanılacak olan eşanlamlılar sözlüğünde kelimeler çekim eklerinden arındırılmış halde bulunduğundan biçimbirimsel belirsizlik gidermeden bir sonraki adım budamadır. Budama işleminde kelimenin aldığı etiketlere göre bazı ekler atılır veya değiştilir. Bu stegosistemde basitlik ve budama kolaylığı açısından isimler, sıfatlar ve zarflar olan üzere 3 farklı kelime tipi üzerine yoğunlaşılmıştır. İsimlerdeki iyelik, hal ve kişi ekleri budanarak 3. tekil kişi yalın hallerine çevrilmiştir. Yapım ekleri ise

Etiketli kelime	**Doğru etiket**
fikir+$Noun+A3pl+Pnon+Gen$	fikir $+Noun+A3pl+Pnon+Gen$
fikir+$Noun+A3pl+P2sg+Nom$	
tart+$Verb+Recip\hat{}DB+Verb+Pass+Neg\hat{}DB+Adj+PastPart+P3sg$	tartış $+Verb\hat{}DB+Verb+Pass+Neg\hat{}DB+Adj+PastPart+P3sg$
tart+$Verb+Recip\hat{}DB+Verb+Pass+Neg\hat{}DB+Noun+PastPart+A3sg+P3sg+Nom$	
tartış+$Verb\hat{}DB+Verb+Pass+Neg\hat{}DB+Adj+PastPart+P3sg$	
tartış+$Verb\hat{}DB+Verb+Pass+Neg\hat{}DB+Noun+PastPart+A3sg+P3sg+Nom$	
yer $+Noun+A3pl+P2sg+Nom$	yer $+Noun+A3pl+P2sg+Nom$
şiddet $+Noun+A3sg+Pnon+Nom$	şiddet $+Noun+A3sg+Pnon+Nom$
devre $+Noun+A3sg+Pnon+Dat$	devre $+Noun+A3sg+Pnon+Dat$
gir $+Verb+Pos+Aor+A3sg$	gir $+Verb+Pos+Aor+A3sg$
gir $+Verb+Pos+Aor\hat{}DB+Adj+Zero$	
. $+Punc$	. $+Punc$

Çizelge 3.3 : Biçimbirimsel belirsizlik gidermeden örnekler.

budanmadan bırakılmıştır. Çizelge 3.4'de etiketli kelimeler ve budanmış halleri birlikte verilmiştir.

3.4 Anlamdaş Arama

Sözcük tabanlı stegosistemler uygulamaya özel sözlüklere ihtiyaç duyarlar çünkü değiştirilen sözcük şüphe uyandırmadan, örtü metninde yer aldığı anlam ve formda örtülü metinde de yer almalıdır. Türkçe'de doğal dil işlemede kullanılabilecek hazır bir eş anlamlılar sözlüğü yoktur. [31]'de yapılan çalışmalar sonucu aynı anlama

Etiketli kelime	**Budanmış kelime**
(elmalarımızda): elma+$Noun+A3sg+P1pl+Loc$	(elma) elma+$Noun+Pnon+Nom$
(devrilebilir): devir+$Verb\hat{}DB+Verb+Pass+Pos\hat{}DB+Verb+Able+Aor+A3sg$	(devrilmek): devir+$Verb\hat{}DB+Verb+Pass+Pos\hat{}DB+Noun+Inf1+A3sg+Pnon+Nom$
(dokunamadığımız): dokun+$Verb\hat{}DB+Verb+Able+Neg\hat{}DB+Adj+PastPart+P1pl$	(dokunmak): dokun+$Verb+Pos\hat{}DB+Noun+Inf1+A3sg+Pnon+Nom$

Çizelge 3.4 : Budama işleminden örnekler.

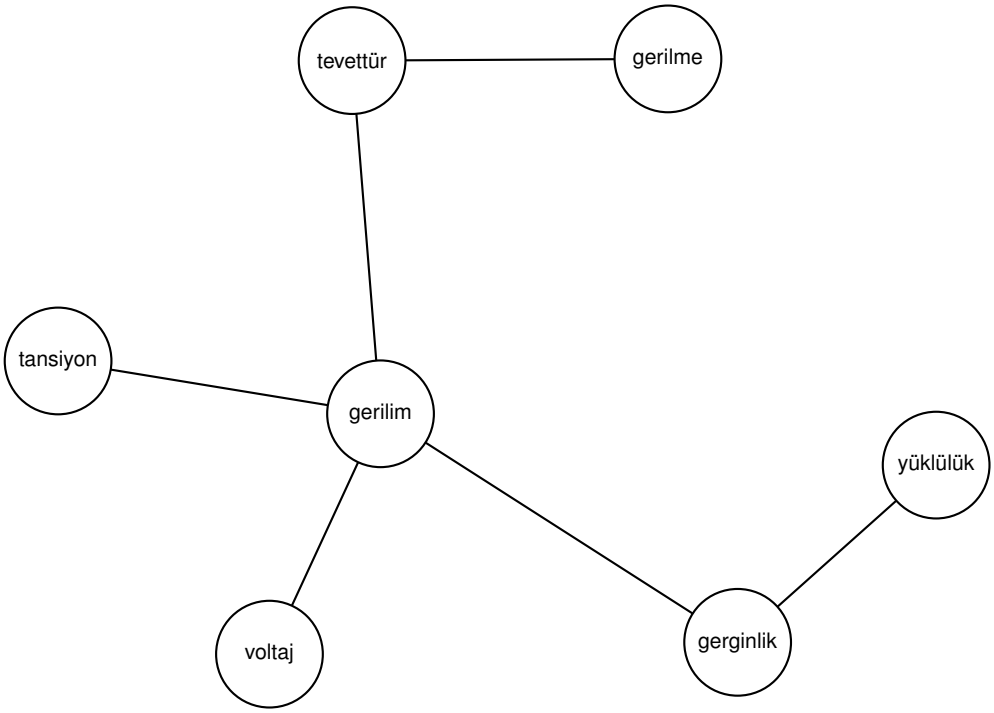
gelen kelimeler eşanlam kümeleri (synsets) haline getirilmiştir. Daha sonra aynı anlama gelmese de yakın anlamları kümelere bu kümeler üzerinden bağlantılar kurularak sözlük genişletilmiştir. Eşanlamlı kümelere erişmek için ise sözlükte bulunan her kelime için hangi eşanlam kümesine bakılması gerektiğini indeksleyen bir yapı tasarlanmıştır. Oluşturulan bu sözlükle küçük denemeler yapılmış ve eş anlamlı kümelerin dağınıklığı, sayılarının yetersizliği ve elle düzenlemelere ihtiyaç duyduğundan güvenlik gerektiren bir stegosistemde kullanılamayacağı anlaşılmıştır. Bu kapsamda tasarlanması gereken ilk bileşen geniş bir eş anlamlılar sözlüğüdür. Kullanılacak olan sözlük anlamlarına ve kelime türlerine göre bir bütün oluşturacak şekilde düzenlenmelidir.

Bir diğer eşanlamlı kelime kaynağı ise Türk Dil Kurumu'nun yaptığı çalışmalardır. 9 Eylül Üniversitesi'nin de yardımıylarıyla oluşan bu kaynakta eş anlamlı kümeler bütünlüğünü korumaktadır ve sistemde gerekli güvenliği sağlayabilir.

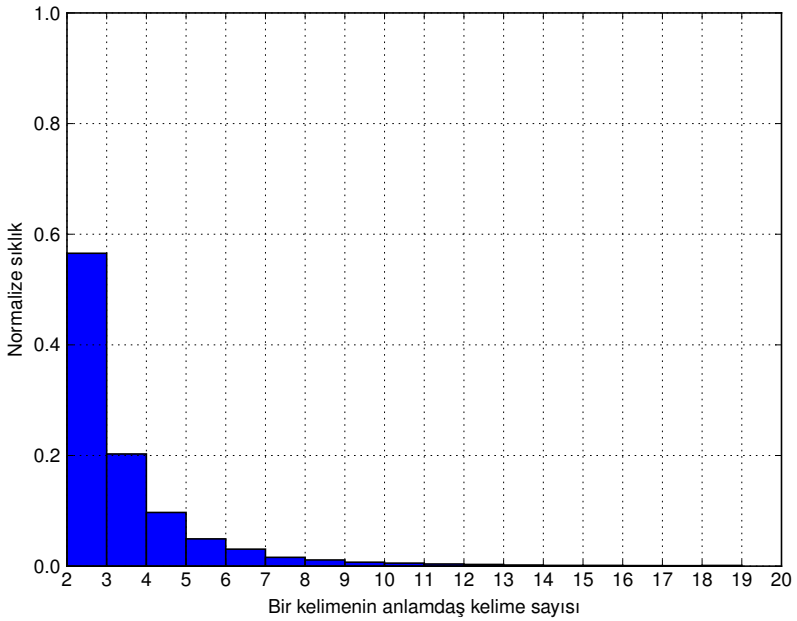
Stegosistemde anlam ilişkilerini kurmak adına sözlükte yer alan bütün kelimeler grafa birer düğüm olarak yerleştirilmiştir. Eş anlama gelenler ise Şekil 3.2' deki gibi ilişkilendirilmiştir. Burada 'gerilim' ile 'voltaj' anlamdaştır. Aynı şekilde 'tansiyon' ile 'gerilim' de anlamdaştır; ama 'tansiyon' ile 'voltaj' anlamdaş değildir. İlerde bu özellikten kodlama aşamasında yararlanılacaktır.

Eş anlamlılar sözlüğünde toplamda 22164 kelime (graph node) ve 47241 ilişki (graph edge) vardır. Kelimelerin eş anlamlı kelimeye sahip olma sayısı ve sıklığı Şekil 3.3'da verilmiştir. Sadece bir adet anlamdaşa sahip olan kelimelerin oranı %58'dir. En fazla 5 anlamdaş kelimeye sahip olanların oranı ise %96'lara kadar çıkmaktadır.

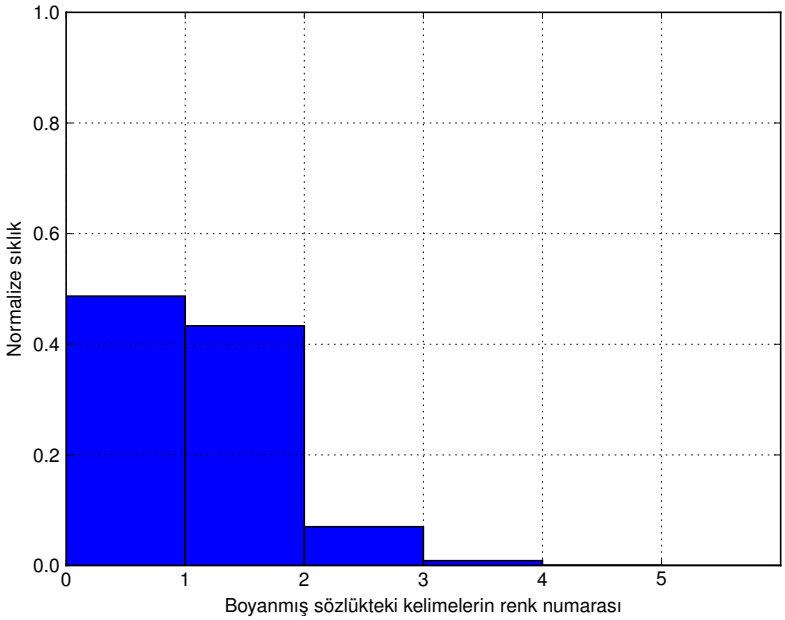
Eş anlamlılar grafından çıkarabileceğimiz bir başka istatitikse grafın kromatik sayısıdır [32]. Graf boyama algoritmasına göre kullanılan eşanlam grafındaki bütün kelimeler 6 farklı renge boyanabilmektedir. 0'dan 5'e kadar numaralandırılmış bu renklerin dağılımı Şekil 3.4'da verilmiştir. Karar kelimesi için boyanmış örnek bir graf ise Şekil 3.5'da verilmiştir. Gönderilen kelimenin veriyi çıkartan tarafta da aynı şekilde çözülebilmesi için boyalı grafta sadece sarı ve yeşil renkli kelimeler aday kelime olarak seçilebilir. Diğer renkli kelimeler gizli veriyi çıkartırken farklı eşanlamlı kelimelere, dolayısıyla da farklı bitlere sebep olacağından bu stegosistemde bilgi taşıyıcı olarak kullanılamazlar.



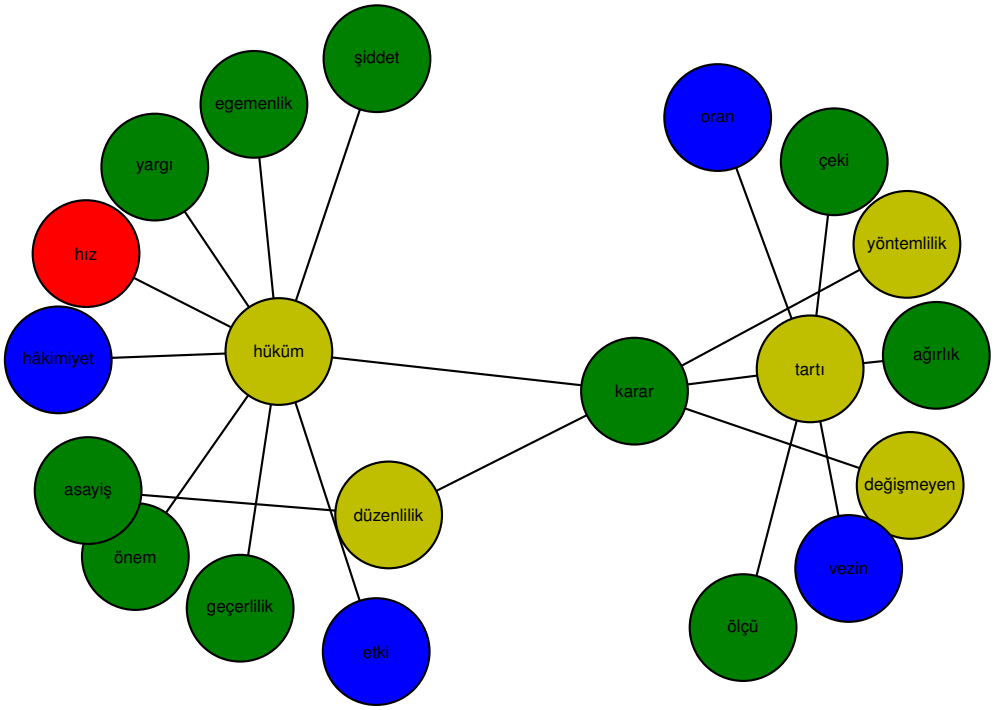
Şekil 3.2 : Eşanlamlılar sözlüğünden oluşturulmuş örnek bir eşanlam grafı.



Şekil 3.3 : Sözlükteki kelimelerin anlamdaş sayıları ve sıklıkları.



Şekil 3.4 : Sözlükteki kelimelerin renk numaraları ve sıklıkları.



Şekil 3.5 : Eşanlamlılar sözlüğünden oluşturulan örnek bir boyanmış graf.

3.5 Puanlama

Eş anlamlılar kümesinden ilgili kelimenin eş anlamları bulunduktan sonra bunların hangisinin bu cümlede daha uygun olduğunu anlamak için aralarında puanlamaya ihtiyaç duyulur. Doğal dil işlemede bir dildeki kelimelerin hangi sıklıkla geçtikleri, hangi kelimelerin beraber kullanıldıkları ve yanyana sıralanma olasılıklarını istatistiksel olarak elde etmek için dil modellemeden yararlanılır. Burada da bu sözcükleri puanlamak için dil modelinden yaralanılacaktır. Tez kapsamında dil modelleme için SRILM yazılımından yararlanılmıştır. [33]. n-gram'lar kullanılarak cümlede değiştirilmek istenen sözcüklerin o cümle özelinde kullanımı araştırılıp belli bir eşikten daha fazla olanlar aday olarak seçilecektir. n adet kelimeden oluşan bir cümledeki kelimelerin yanyana gelme olasılığının hesabı denklem 3.1'deki gibi yapılır:

$$P(w_1 w_2..w_n) = \prod_{i=1}^{n} P(w_i | w_1 w_2..w_{i-1}) \tag{3.1}$$

3-gramlar kullanılarak oluşturulan bu modelde her zaman orjinal değişmemiş cümle içinden bir kelime seçilir, bu kelimeyle aynı anlama gelen bütün anlamdaş kelimeler sırayla orjinal cümleye yerleştirilir ve cümle tekrar puanlanır. Çizelge 3.5'de "tartışmada gerilim giderek arttı" örnek cümlesindeki 'gerilim' aday kelimesinin eşanlamlılarının cümleye yerleştirilmesiyle oluşan yeni cümleler ve dil modelinden gelen puanları hesaplanmıştır. Bu örnekte cümleye istatistiksel olarak en uygun kelime 'giderek' kelimesiyken en uzak kelime ise 'anbean' kelimesidir.

Cümle	Puanı
tartışmada gerilim giderek arttı	-13.14
tartışmada gerilim tedricen arttı	-19.01
tartışmada gerilim anbean arttı	-20.21
tartışmada gerilim gitgide arttı	-18.45

Çizelge 3.5 : Dil modelini kullanılanarak puanlanan örnek cümleler.

Dili modellemek için kullanılan olan veri Sak'ın hazırladığı derlemlerdir [34]. Bu derlemlerden biri 491175 adet farklı gazete haberlerinden, diğeri de 489874 adet farklı genel amaçlı metinlerden meydana gelmiştir. Cümledeki kelimeler bu aşamada budanmış olduğundan dil modeli de budanmış cümlelerden oluşturulmuştur.

3.6 Tekrar Oluşturma

Sözlükte ilgili kelimenin eş anlamlısını bulmak için kelimeler Bölüm 3.3'de verilen budama işleminden geçmişti. Bu işlemin tersi olarak da yeni kelimeyi cümlede geçtiği formata geri döndürmek için tekrar ouşturma işlemi uygulanır; diğer bir ifadeyle kelime kişi, iyelik ve hal eklerini geri alarak cümleye uyum sağlayacak şekilde tekrar oluşturulur. Çizelge 3.6'de budanmış kelimeler ve tekrar oluşturulmuş halleri birlikte verilmiştir.

Budanmış kelime	**Tekrar oluşturulmuş kelime**
(armut) armut+$Noun + Pnon + Nom$	(armutlarımızda): armut+$Noun + A3sg + P1pl + Loc$
(düşmek): devir+$Verb\^{}DB + Verb + Pass + Pos\^{}DB + Noun + Inf1 + A3sg + Pnon + Nom$	(düşebilir): devir+$Verb\^{}DB + Verb + Pass + Pos\^{}DB + Verb + Able + Aor + A3sg$
(dokunmak): dokun+$Verb + Pos\^{}DB + Noun + Inf1 + A3sg + Pnon + Nom$	(ilişemediğimiz): iliş+$Verb\^{}DB + Verb + Able + Neg\^{}DB + Adj + PastPart + P1pl$

Çizelge 3.6 : Tekrar oluşturma işleminden örnekler.

3.7 Kodlama

Bu bölümde örtü metnine gizli bilgiyi gömen tarafla örtülü metindeki gizli bilgiyi çıkartan taraf arasında hem senkronizasyonu sağlamak, hem de üçüncü kişilerce bilgi taşıyıcı olan kelimelerin temsil ettiği bit değerlerini rastgeleleştirmek amacıyla kaotik bir sistemden yararlanılarak tasarlanan kodlayıcı anlatılacaktır. Başlangıç değeri olan x_0 değerinin ve denklem parametresi olan k değerinin haberleşmede gizli anahtar olarak kullanıldığı kaotik sistemin entropisi ve dağılımı incelenerek tasarlanan stegosistemde kodlayıcı olarak nasıl rol aldığı anlatılacaktır.

3.7.1 Koatik Haritalar

Koatik haritalar ayrık zamanlı ve kaotik davranış gösteren dinamik haritalardır. Denklem 3.2'de genel bir kaotik sistemin yapısı verilmiştir:

$$\boldsymbol{x}_{n+1} = \boldsymbol{f}(\boldsymbol{x}_n); \quad \boldsymbol{x} \in S \subseteq R^N; \quad \boldsymbol{f} : S \to S \tag{3.2}$$

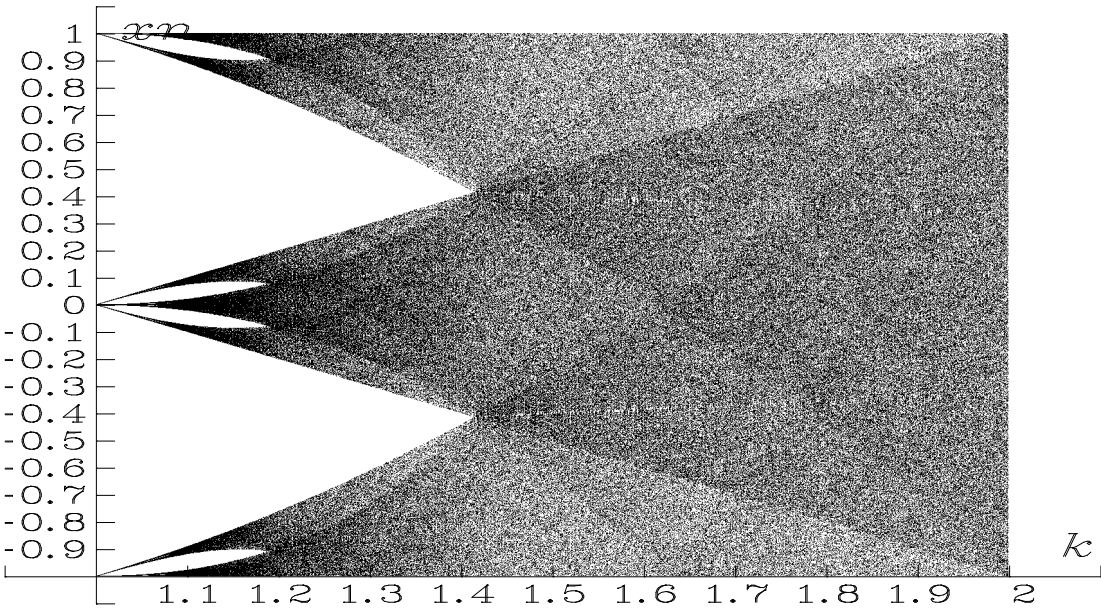
Kaotik sistemlerin başlangıç koşullarına aşırı duyarlılığı, gürültü benzeri spektral güç yoğunluğuna (PSD) sahip olmaları ve düzensiz ve aperiyoduk davranışları rastgelelik gerektiren uygulamalarda entropi kaynağı olarak kullanılmalarına olanak sağlar [35].

3.7.2 PL1D

Tek boyutta parçalı lineerlik gösteren PL1D (Piecewise Linear 1 Dimensional Map) haritaları ayrık zamanlı kaotik haritaların bir alt kümesidir. Literatürdeki en bilinen örnekleri Bernoulli Shift, Tent Map ve Skew Tent Map'tir. Basitliği, uygulanabilirliği ve ispatlanabilirliği PL1D haritaların rastegele sayı üretmek için kullanımında başı çeken özelliklerindendir [36]. Denklem 3.3 önerilen stegosistemde gerekli entropiyi sağlamak üzere kullanılacak PL1D haritasıdır.

$$x_{n+1} = f(x_n) = \begin{cases} kx_n - 1, & \text{for } x_n \geq 0 \\ kx_n + 1, & \text{for } x_n < 0 \end{cases} \qquad (3.3)$$
$$x_n \in (-1,1) \quad k \in (1,2)$$

Denklemdeki tek parametre olan k sistemdeki kaotik davranışı dinamik olarak sağlar. Şekil 3.6'deki dallanma diyagramı'ndan da (bifurcation diagram) görüldüğü gibi:



Şekil 3.6 : Önerilen kaotik haritanın dallanma diyagramı.

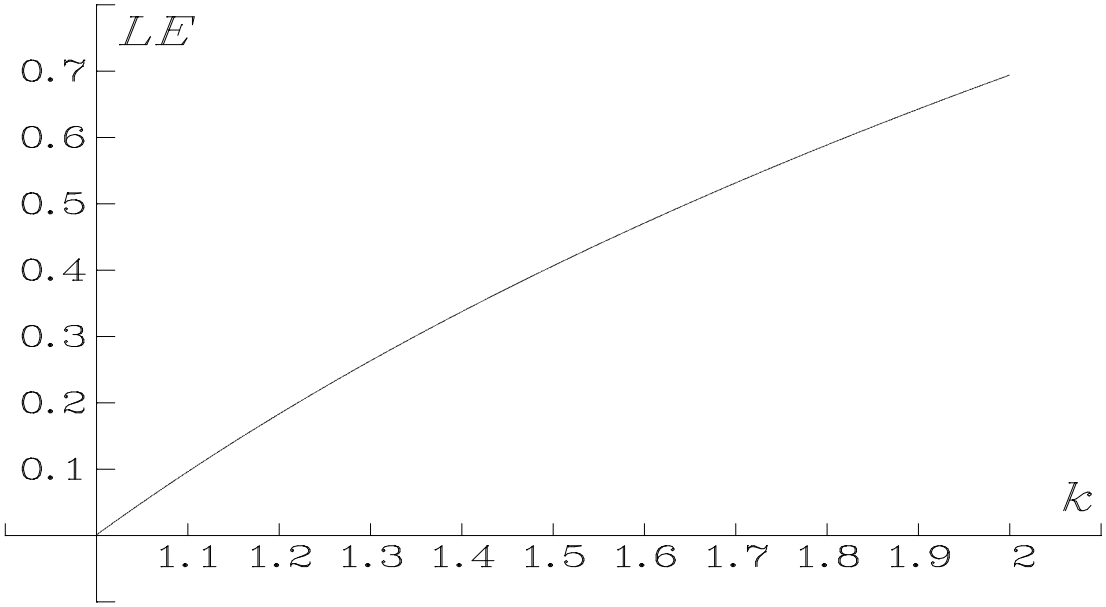
(i) $1 < k \leq \sqrt[4]{2}$ aralığında 7 ve daha fazla ayrık x_n kümeleri,

(ii) $\sqrt[4]{2} < k \leq \sqrt{2}$ aralığında 3 farklı x_n kümeleri,

(iii) $\sqrt{2} < k \leq 2$ aralığında tek bir x_n kümesi bulunur [36].

Önerilen haritanın Lyapunov üsteli (Lyapunov exponent) grafiği denklem 3.4'te verilen ifade yardımıyla hesaplanmış ve Şekil 3.7'de verilmiştir.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \ln |f'(x_i)| \tag{3.4}$$



Şekil 3.7 : Önerilen kaotik haritanın Lyapunov üsteli diyagramı.

Açıkca görüldüğü gibi $k > 1$ için Lyapunov üsteli her zaman pozitiftir. Diğer bir ifadeyle dinamik sistem bu aralıkta kaotik davranış gösterir ve sistemde entropi kaynağı olarak kullanmaya uygundur [35].

3.8 Kaotik Kodlayıcı

Önerilen sistemin kaotik davranışı dallanma diyagramı ve Lyapunov üsteli grafiğiyle gözler önüne serildikten sonra bu bölümde kaotik kodlayıcının tasarımına yer verilecektir. Sistemde o anda üzerinde çalışılan cümle için ($sentence_i$), boyanmış grafta hangi rengin hangi biti temsil ettiğini anlamak adına bir bitlik lojik bir ifadeye (bit_i), aday kelimelerden ($word_{ij}$) hangisinin seçildiğini belli etmek için ise bir adet tamsayı ifadeye (int_i) ve o cümle içindeki toplam aday kelime sayısına ihtiyaç vardır. Aday kelimeler puanı orjinal cümledeki kelimenin t komşuluğundaki kelimelerdir. Denklem 3.5 i. cümle için bu iki değişkenin nasıl hesaplanacağını ortaya koyar:

$$
\begin{aligned}
bit_i &= sgn(x_i) \\
int_i &= ((x_i \mod 1) >> 32) \mod (\sum_{j=1}^{m} c_{ij}) \\
c_{ij} &= \begin{cases} 1, & \text{for } |score(word_{ij}) - score(sentence_i)| \leq t \\ 0, & otherwise \end{cases}
\end{aligned}
\tag{3.5}
$$

Bu denkleme göre bit_i değeri signum fonskiyonun da anlaşıldığı gibi kaotik değişken olan x_i'nin o indisli işaretine bağlıdır. int_i'nin değeri ise o indisli kaotik değişkenin virgülden sonraki kısmının toplam aday sayısına göre mod alınmasıyla oluşur.

i. cümledeki bilgi yüklenecek kelimenin seçimi için önce cümledeki toplam aday kelime sayısı bulunur. i. cümledeki j. kelimenin ($word_{ij}$) aday cümle olup olmadığını anlamak için ise sözlükte o kelimenin eş anlamlısı cümleye yerleştilir ve dil modeli yardımıyla cümle puanlaması tekrar yapılır. Orjinal cümleyle eş anlamlısıyla değiştirilmiş cümle arasındaki fark sistemde eşik değeri olan t değerinden küçükse bu kelime bu cümle için bir aday kelimedir ($c_{ij} = 1$) denir. Aksi halde kelimenin aday kelime olmadığı anlaşılır ve cümledeki bir sonraki kelimeye geçilir. Cümlede hiç bir aday kelime bulunamıyorsa cümleye herhangi bir veri gömülemeyeceği anlaşılır ve bir sonraki cümleye geçilerek işlem tamamlanır. Algoritma 1 ile metinde geçen cümlelerden hangilerine bilgi yüklenebileceğini, yüklenen bu bilginin hangi kelimede olacağını ve ilgili renk indisinin hangi lojik ifadeyke kodlanacağını kararlaştıran kod parçası verilmiştir.

Algoritma 1 Bilgi taşıyıcı kelimelerin bulunması ve kodlanması.

for $i = 1$ to n **do**
  $candidate_i = \emptyset$
  $x_i = f(x_{i-1})$
  for $j = 1$ to m **do**
    $synset_{ij} = \emptyset$
    $synonyms = synonym(word_{ij})$
    for $word_{syn} \in synonyms$ **do**
      if $color(word_{syn}) \in \{yellow, green\}$ **then**
        if $|score(word_{syn}) - score(word_{ij})| \leq t$ **then**
          $synset_{ij} = synset_{ij} \cup \{word_{syn}\}$
        end if
      end if
    end for
    if $|synset_{ij}| > 0$ **then**
      $candidate_i = candidate_i \cup \{j\}$
    end if
  end for
  l = $|candidate_i|$
  if $l > 0$ **then**
    $int_i = ((x_i \mod 1) >> 32) \mod l$
    $bit_i = sgn(x_i)$
  end if
end for

4. ÖNERİLEN STEGOSİSTEMLE GİZLİ VERİ İLETİŞİMİ

Tasarlanan stegosistem ve alt bileşenleri Bölüm 3'de ayrıntılarıyla verildikten sonra bu bölümde tasarlanan bu sistemle gizli verinin gönderilen tarafta nasıl gömüleceği ve alıcı tarafta nasıl çözüleceği örneklerle anlatılacaktır.

4.1 Veri Gömme

Örtü metnine gizli bilginin gömülmesini anlatan adımlar Algoritma 2 ile verilmiştir.

Algoritma 2 Gizli verinin örtü metnine gömülmesi.

1: $index_{secret} = 1$
2: **for** $i = 1$ to n **do**
3: $secretbit = secret[index_{secret}]$
4: l = $|candidate_i|$
5: **if** $l > 0$ **then**
6: $j = int_i$
7: **if** $color(word_{ij}) = green$ **then**
8: $currentbit = 0$
9: **else**
10: $currentbit = 1$
11: **end if**
12: **if** $secretbit \neq currentbit \oplus bit_i$ **then**
13: $word_{ij} = bestscored(synset_{ij})$
14: $regenerate(sentence_i)$
15: **end if**
16: $index_{secret} = index_{secret} + 1$
17: **end if**
18: **end for**

Algoritma 2 daha önce Algoritma 1 ile bulunan $candidate_i$, int_i, bit_i, $synset_{ij}$ gibi değişkenler yardımıyla örtü metninin gizli bilgiyi de içerek şekilde örtülü metne evrilmesini sağlar. 5. satırda cümlenin bilgi gömülebilen aday bir cümle olup olmadığına bakılır. Bilgi gömülemeyecek bir cümle ise bu cümle pas geçilir, değilse 6. satırda bilginin gömüleceği kelime int_i değişkeni yardımıyla seçilir ve 7-12 satırları arasında kelimenin temsil ettiği bit değeri bulunur. 12. satırda gömülmek istenen gizli veriyle metinde hali hazırda bulunan kelimenin temsil ettiği bit değerine bakılır.

Burada bit_i değişkeni hem gönderici, hem de alıcı tarafta senkronizasyonu sağlayarak üçüncü kişilerce gizli bilginin kestirimini engeller çünkü her bir cümlede kaotik kodlayıcıdan gelen bit_i değerine göre kelimenin temsil ettiği bit değeri değişir. Eğer orjinal kelime istenilen bit değerini temsil ediyorsa herhangi bir değişikliğe ihtiyaç yoktur, veri zaten hazırdır. Eğer temsil etmiyorsa anlamdaş kelimelerden en iyi puanlısı seçilerek cümle tekrar oluşturulur ve bir sonraki cümleye geçilir.

4.2 Veri Çıkartma

Örtülü metinden gizli bilginin çıkartılmasını anlatan adımlar Algoritma 3 ile verilmiştir.

Algoritma 3 Gizli verinin örtülü metinden çıkartılması.

1: $index_{secret} = 1$
2: **for** $i = 1$ to n **do**
3: l = $|candidate_i|$
4: **if** $l > 0$ **then**
5: $j = int_i$
6: **if** $color(word_{ij}) = green$ **then**
7: $currentbit = 0$
8: **else**
9: $currentbit = 1$
10: **end if**
11: $secretbit = currentbit \oplus bit_i$
12: $secret[index_{secret}] = secretbit$
13: $index_{secret} = index_{secret} + 1$
14: **end if**
15: **end for**

Algoritma 3 daha önce Algoritma 1 ile bulunan $candidate_i$, int_i, bit_i gibi değişkenler yardımıyla örtülü metindeki gizli bilgiyi bit bit okumayı sağlar. 5. satırda cümlenin bilgi gömülebilen aday bir cümle olup olmadığına bakılır. Bilgi gömülemeyecek bir cümle ise gönderici tarafından da bu cümleye herhangi bir veri gömemeyeceği anlaşılır ve cümle pas geçilir, değilse 6. satırda bilginin gömüleceği kelime int_i değişkeni yardımıyla seçilir ve 7-12 satırları arasında kelimenin temsil ettiği bit değeri bulunur. Gizli bilgi ise bu bit değerinin bilgiyi gömen tarafta da yapıldığı gibi bit_i değeri ile $\oplus$ (XOR) işleminden geçirilmiş halidir.

5. ÖNERİLEN STEGOSİSTEMİN DEĞERLENDİRMESİ

Dilbilimel sistemler genellikle iki farklı şekilde değerlendirilir. Bunlardan ilki bilgisayar yardımıyla hesaplanabilen veri gömme kapasitesi, dayanıklılık ve güvenlik gibi steganografide kullanılan temel metriklerin teorik ve istatistiksel olarak ifade edilmesidir. Diğeri ise insanlar tarafından değerlendirilen son kullanıcı görüşleridir. Tasarlanan stegosistemlerin direkt olarak değerlendirilmesi olanak sağlayan bu ikinci metotta örtülü metinler kullanıcıların karşısına çıkartılır ve cümle doğallığı, cümleyi doğal haline getirmek için yapılması gereken minimum değişiklik gibi parametrelerle sistemin başarısı kestirilir.

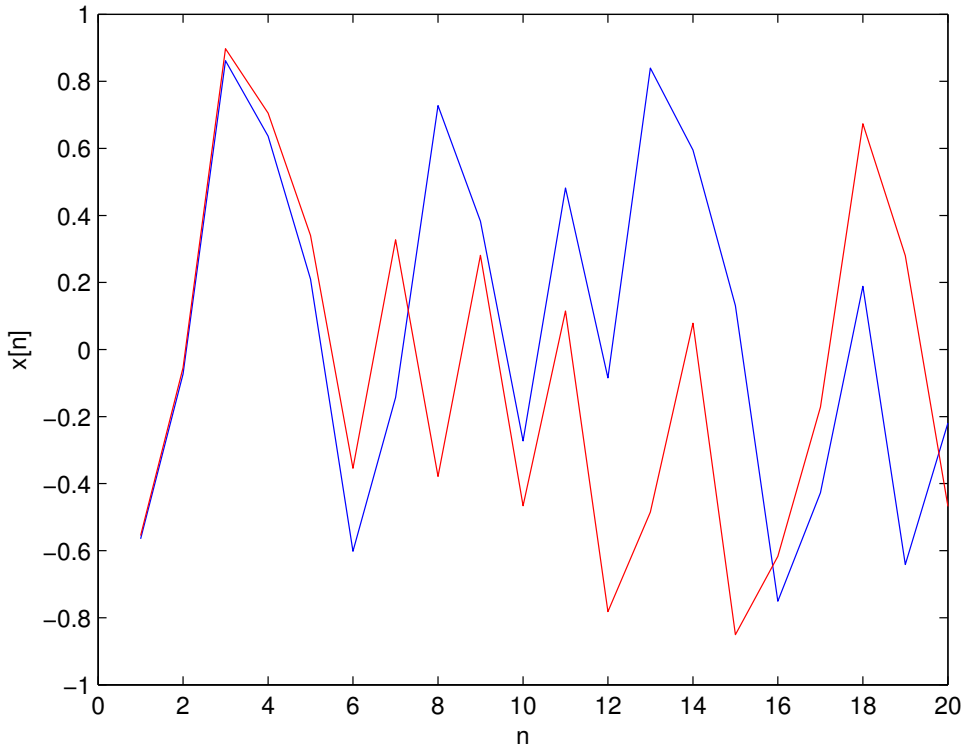
5.1 Teorik Değerlendirme

5.1.1 Güvenlik

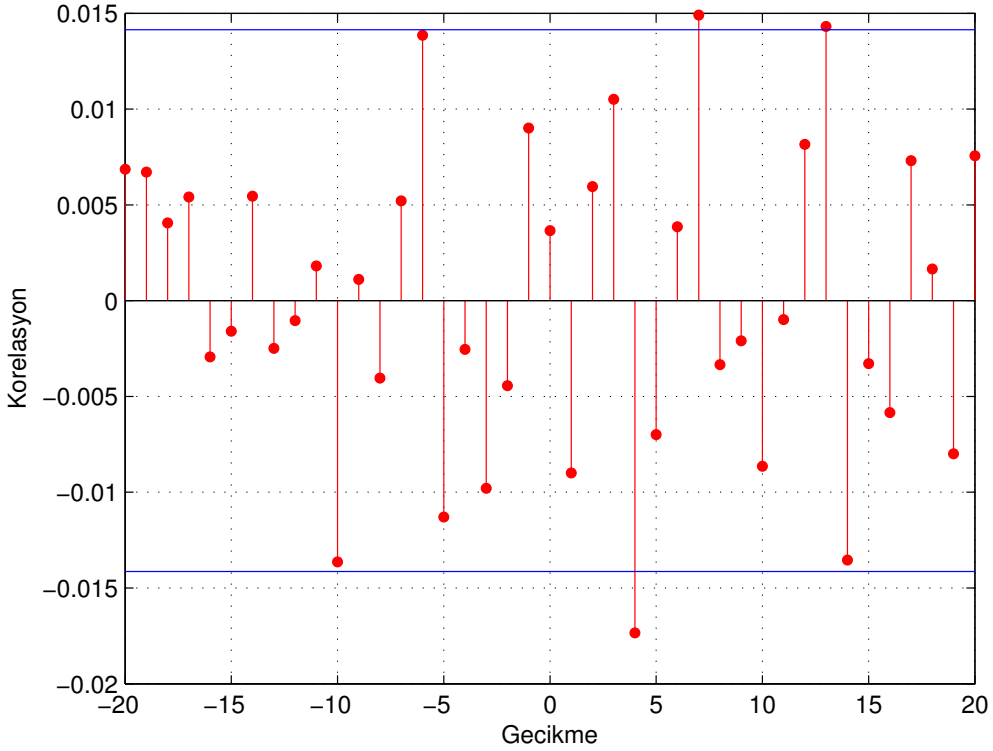
Gizli anahtarlardaki en küçük sapmanın bütün senkronizasyonu bozarak üçüncü kişilere karşı sağladığı güvenlik aşağıda şekiller üzerinde gösterilmiştir. Şekil 5.1'de kaotik denklemin 0.01 yakınlıktaki iki farklı başlangıç değeri x_0 için alınmış x_n değerleri çizdirilmiştir. Görüldüğü gibi n>6 için diziler birbirinden hızla ıraksamaktadır. Dizilerden 20000 uzunluklu kesitler alınmış ve korelasyon katsayıları (sample cross correlation) Şekil 5.2'de görüldüğü gibi bulunmuştur. İki dizi arasındaki korelasyonun %95'lik dilimde kaldığına dikkat ediniz.

Şekil 5.3 ve 5.4'de ise bir önceki grafikte x_0 için yapılanlar burada sistem parametresi olan k için tekrarlanmıştır. Burada da ilintisiz dizilerin birbirlerinden hızla ıraksadıkları açıkça görülmektedir. Özetle gönderici ve alıcı tarafların gizli kanal üzerinden paylaştıkları bu iki gizli anahtarla stegosistemin güvenliği sağlanır. Kerckhoff prensibinde de açıkça belirtildiği gibi stegosistem tümüyle bilinse bile gizli anahtarlar sistemi güvende tutar [37].

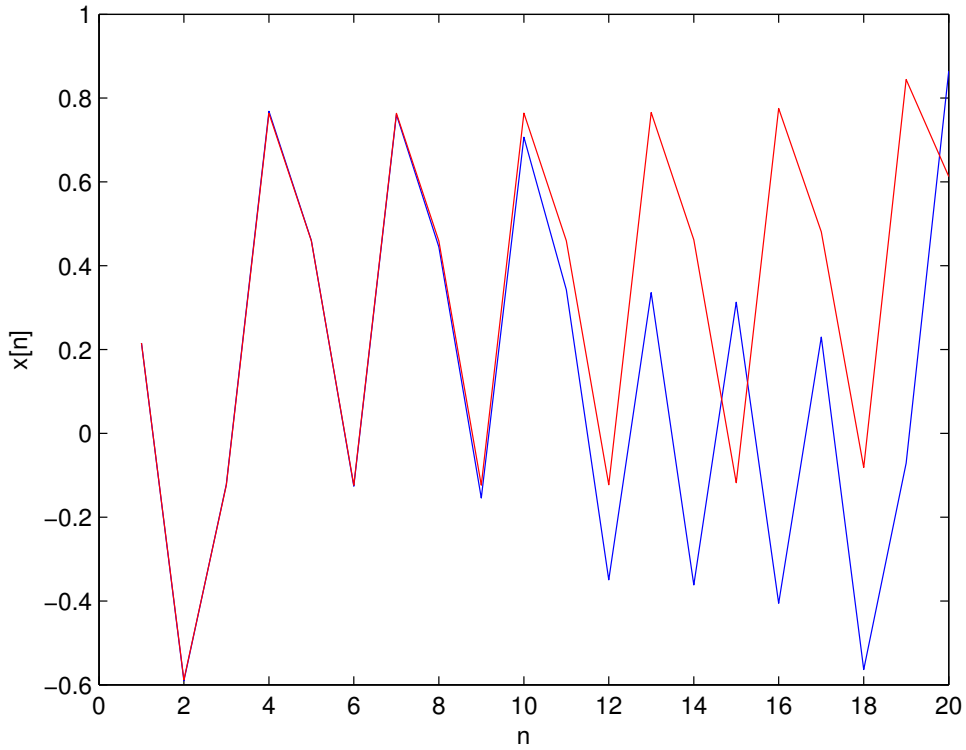
5.1.2 Kapasite



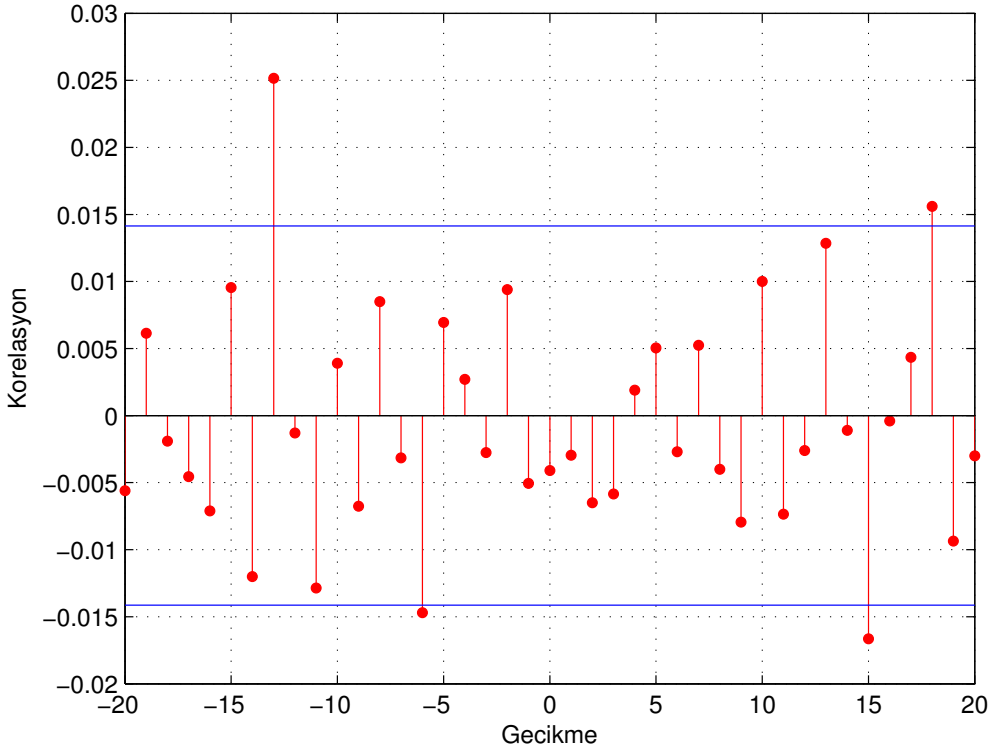
Şekil 5.1 : Başlangıç değeri olan x_0 değerinin 0.01 sapması ile oluşan iki farklı x_n dizisi



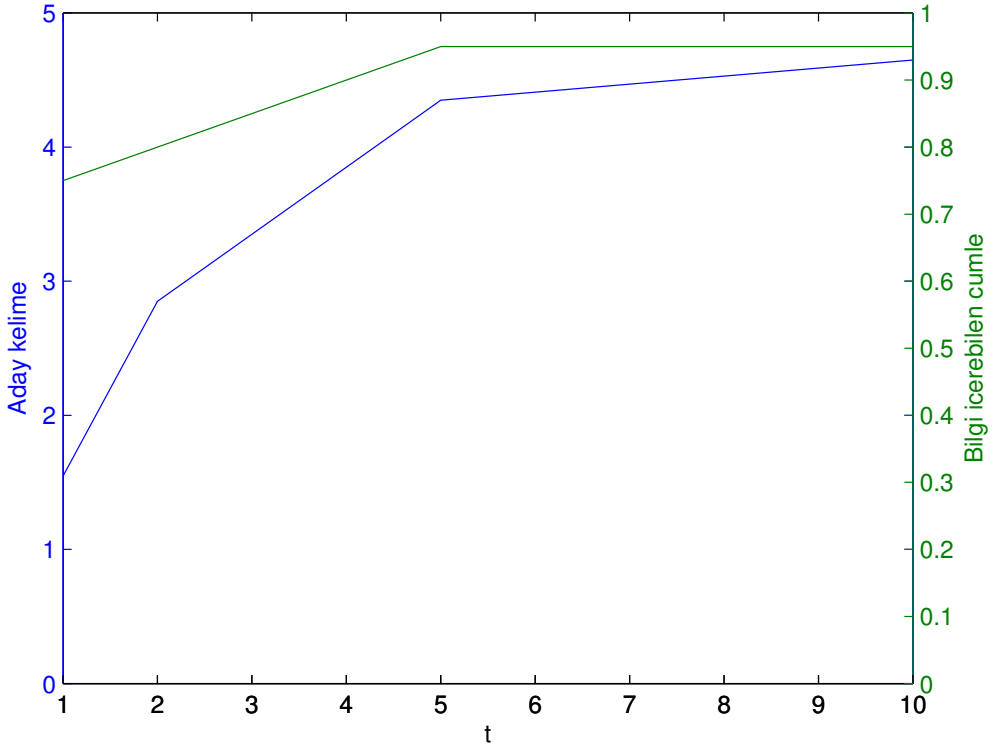
Şekil 5.2 : Başlangıç değeri olan x_0 değerinin 0.01 sapması ile oluşan iki farklı x_n dizisinin korelasyon grafiği



Şekil 5.3 : Denklem parametresi olan k değerinin 0.01 sapması ile oluşan iki farklı x_n dizisi.



Şekil 5.4 : Denklem parametresi olan k değerinin 0.01 sapması ile oluşan iki farklı x_n dizisinin korelasyon grafiği.



Şekil 5.5 : Farklı eşik değerlerine karşılık bulunan normalize kapasite.

Stegosistemin kapasitesi cümle başına gizlenebilen bit sayısıyla ifade edilmiştir (bit/cümle). Eşik değeriyle (t) doğru orantılı olarak değişir. t değerinin sırasıyla 1, 2, 5 ve 10 değerleri için örnek olarak seçilmiş 20 cümledeki ortalama aday kelime sayısı ve bilgi içeren cümle sayısının grafikleri Şekil 5.5'de verilmiştir. Önceki bölümlerde de anlatıldığı gibi cümlede en az bir adet bilgi taşıyıcı aday kelimenin olması o cümleye bilgi yüklemek için yeterli olduğundan bilgi içerebilen cümle sayısı direkt olarak kapasitenin bir ölçüsüdür. Türkçe için yapılan sentaktik tabanlı çalışmaya göre bu oran istenen güvenliğine göre ayarlanabilen t parametresiyle yaklaşık 8 kat artırılmıştır.

Dikkat edilmesi gereken diğer bir husus ise eşiğin belli bir değerden sonra kapasiteyi artırmadığına dikkat ediniz. Bunun sebebi ilgili graftaki aday anlamdaş kelimelerin neredeyse tümünün kapsanmasıdır. Dil modelindeki puanlama logaritmik ölçekte yapıldığından $t > 5$ için artık kelimeler birbirinden hızla uzaklaşmaya başlar.

5.1.3 Dayanıklılık

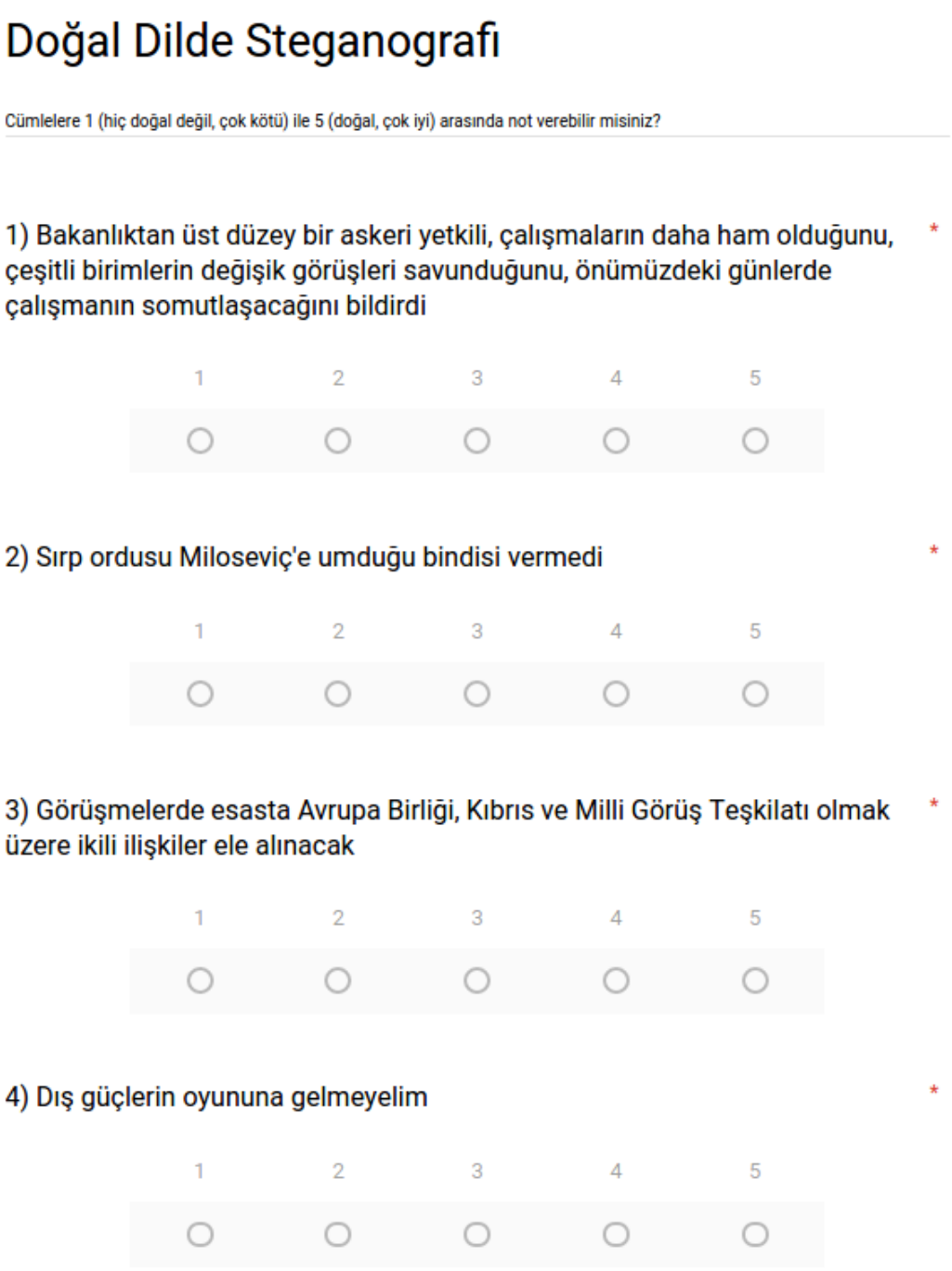
Stegosistemde cümle başına yapılan değişiklik en fazla bir kelimede olmaktadır ve bu kelime tamamen gizli anahtarlarla kaotik kodlayıcıda üretilen int_i değerinde bağlıdır. Damgalamanın aksine steganografide üçüncü kişilerin aktif değil pasif kişiler olduğu

göz önünde bulundurularak örtülü metinde herhangi bir bozulmaya (forgery) neden olmayacağı kabul edilir. Tasarlanan bu sistem bu kabul yapılmasa bile rakiplerine göre avantajlar sunar çünkü hangi kelimede değişiklik yapıldığını sadece gizli anahtara sahip olan kişiler bilir. i. cümlede m_i farklı bilgi yüklenebilen kelime olması halinde aktif bir üçüncü kişinin değiştirilen kelimeyi tahmin etme olasılığı $1/m_i$'dir. Bu kelimenin istenilen bit değerini zaten temsil etme olasılığı $(1/2)$ da göz önüne alınırsa bu oran $1/2m_i$ değerine düşer. Metin bazında düşünüldüğünde ise üçüncü bir kişinin gizli bilgileri tahmin etme olasılığı bilgi yüklenebilecek bütün cümleler için aynı oranın çarpılması kadar azalır. Eşik değeri olan t değerinin artmasıyla bu oran azalır ve sistem gizli bilgiyi üzerinde tutma anlamında daha dayanıklı hale gelir. Tabiki bu eşik yükselmesinin güvenlik açıkları doğurabileceği unutulmamalıdır.

5.2 Pratik Değerlendirme

Örtülü veriyi örtülü kanalda ortaya daha iyi çıkarabilecek olan insan olduğu için sistem değerlendirmesinin diğer çalışmalarda da olduğu gibi insan gözüyle değerlendirilmesi burada da uygun görülmüştür. [25] [38] [28]. Sistemi değerlendirmek üzere gazete haberlerinden rastgele olarak alınan 20 cümlelik veri kümesi derlenmiştir. Bu cümleler eşik değerinin sırasıyla 1, 2, 5 ve 10 değerleri için stegosistemden geçirilmiş ve hangi cümle üzerinde ne değişiklik yapıldığı, hangi cümlenin değşiklik gerektirmeden bilgi içerdiği ve hangi cümlenin bilgi içeremediği kaydedilmiştir. Kullanıcılara ise bu konuda herhangi bir bilgi verilmemiştir. Daha sonra değerlendirilmeleri için oluşan bu 80 cümle anadili Türkçe olan 21 farklı kullanıcıya rastgele olarak sunulmuştur. Kullanıcılardan bu cümleleri 1 ile 5 arasında en doğal olanı 5, en göze batanı 1 olacak şekilde doğallığına göre notlandırmaları istenmiştir. Anket sayfasından alınmış örnek bir ekran görüntüsü Şekil 5.6'de verilmiştir.

Kullanıcıların puanlamasına göre çıkan sonuçlar ise Çizelge 5.1'deki gibidir: İlk sütün eşik değeri olan t parametresidir. 2. ve 3. sütünlarda ise eşik değerinin 1, 2, 5 ve 10 değeri için sırasıyla 7, 7, 12 ve 12 değişiklik gerektiren her bir cümlenin kişilere göre ortalama değerinin ortalaması ve standart sapması verilmiştir. Eşik değerinin arttıkça ortalamanın azalmasına ve standart sapmanın artmasına dikkat ediniz.



Şekil 5.6 : Son kullanıcılara yapılan anketten örnek bir ekran.

Eşik	Ortalama	Standart sapma
1	3.43	0.35
2	3.33	0.39
5	3.18	0.68
10	3.16	0.72

Çizelge 5.1 : Değiştilmiş cümlelerdeki ortalama puanlar ve standart sapmaları.

Üzerinde hiç değişiklik yapılmamış cümlelerde ortalama değer 3.78, standart sapma ise 0.40 çıkmaktadır. Bu da doğal dildeki yazım tarzı ve kelime seçimi gibi seçeneklerin steganografide fazladan bir güvenlik marjı yaratmaktadır.

6. SONUÇLAR VE ÖNERİLER

Çağımızda güvenlik her alanda ihtiyaç duyulan bir teknoloji haline gelmiştir ve bilginin gizlili artan iletişim hızları ve işlem güçleriyle birlikte önemini gitgide artırmaktadır. Klasik kriptoloji iletişim yapıldığı gerçeğini gizlememekle mahremiyet sorunlarını beraberinde getirmektedir. Steganografi ise buradaki potansiyel riskleri bertaraf ederek gerekli güvenliği ve gizliliği sağlar. Doğal dil ve metinler hala üzerinde en çok çalışan iletişim ortamı olma özelliğini koruyarak steganografi için cazip bir ortam sunmaktadır.

6.1 Sonuçlar

Bu tez kapsamında Türkçe dilinde kelime tabanlı bir stegosistem tasarlanmıştır. Varolan çözümler incelenmiş, Türkçe için daha önceden kelime tabanlı bir steganografi uygulması yer almadığı görülmüş ve çalışmalar bu yönde ilerletilmiştir. Türkçe için anlam ve biçim ilişkisi gözönünde bulundurularak sistematik bir sözlük oluşturulmuş ve örtü metninde bilgi taşıyıcısı olarak kullanılan anlamdaş sözcüklerin alternatiflerini bulmak için kullanılmıştır.

Kelimeler bir grafa dökülmüş ve anlamdaş kelimeler aralarında ilişki kurularak iletilmek istenen bite göre dil modelinden en yüksek puanı alan eş anlamlı kelimeye yer verilmiştir.

Türkçedeki daha önce yapılan sentantik tabanlı çalışmaya göre veri gömme oranı 5 ile 8 kat artırılmıştır. Sistemdeki güvenlik parametresi de güvenlik-veri gömme oranı ödünleşimini (trade-off) istenilen yönde kullanmaya yardım etmektedir.

Literatüre yapılan bir diğer katkı da kaotik bir kodlayıcıdır. Varolan sistemlerdeki güvenlik açıklarını da kapatmak üzere ayrık yapılı kaotik haritalar kullanılarak sistemdeki gerekli entropi sağlanmıştır. Veriyi gömen ve çıkartan taraflar arasında senkronizasyonu sağlayan kaotik haritanın başlangıç değeri, kazanç katsayısı ve eş

anlamı kelimeleri belirlemedeki eşik ise kapalı anahtarlar olarak sistemdeki yerini almakta ve istenilen güvenliği sağlamaktadır.

6.2 Öneriler

Yapılan bu çalışmanın ileride daha geniş bir sözlükle ve daha kapsamlı bir budama mekanizmasıyla çok daha geniş alternatif kümeler oluşturabilen bir stegosistem olması hedeflenmektedir. Wordnet benzeri yapılarla kelimeler arasındaki anlam ilişkileri kurularak sistem güçlendirilecek ve güvenliği artırılacaktır. Anlam belirsizliğini giderme (WSD) araçları sisteme entegre edilerek çok daha mantıklı ve göze batmayan metinler oluşturulması amaçlanmaktadır.

İncelenmeye değer diğer bir husus ise tek bir dil modeli yerine örtü metnine göre değişen bir dil modeli kullanmaktır. Böylelikle örtülü metniyle çok daha alakalı alternatif kelimelerin bulunması hedeflenmektedir. istatistiksel olarak

KAYNAKLAR

[1] **Simmons, G.J.** (1984). The subliminal channel and digital signatures, *Workshop on the Theory and Application of of Cryptographic Techniques*, Springer, s.364–378.

[2] **Simmons, G.J.** (1984). The prisoners' problem and the subliminal channel, *Advances in Cryptology*, Springer, s.51–67.

[3] **Wayner, P.** (1992). Mimic functions, *Cryptologia*, *16*(3), 193–214.

[4] **Wayner, P.** (1995). Strong Theoretical Stegnography, *Cryptologia*, *19*(3), 285–299.

[5] **Anderson, R.** (1996). Stretching the limits of steganography, *Information Hiding*, Springer, s.39–48.

[6] **Chapman, M. ve Davida, G.** (1997). Hiding the hidden: A software system for concealing ciphertext as innocuous text, *International Conference on Information and Communications Security*, Springer, s.335–345.

[7] **Chapman, M.**, **Davida, G.I. ve Rennhard, M.** (2001). A practical and effective approach to large-scale automated linguistic steganography, *International Conference on Information Security*, Springer, s.156–165.

[8] **Cachin, C.** (1998). An information-theoretic model for steganography, *International Workshop on Information Hiding*, Springer, s.306–318.

[9] **Petitcolas, F.A.**, **Anderson, R.J. ve Kuhn, M.G.** (1999). Information hiding-a survey, *Proceedings of the IEEE*, *87*(7), 1062–1078.

[10] **Atallah, M.J.**, **Raskin, V.**, **Crogan, M.**, **Hempelmann, C.**, **Kerschbaum, F.**, **Mohamed, D. ve Naik, S.** (2001). Natural language watermarking: Design, analysis, and a proof-of-concept implementation, *International Workshop on Information Hiding*, Springer, s.185–200.

[11] **Atallah, M.J.**, **Raskin, V.**, **Hempelmann, C.F.**, **Karahan, M.**, **Sion, R.**, **Topkara, U. ve Triezenberg, K.E.** (2002). Natural language watermarking and tamperproofing, *International Workshop on Information Hiding*, Springer, s.196–212.

[12] **Bennett, K.** (2004). Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text.

[13] **Grothoff, C.**, **Grothoff, K.**, **Alkhutova, L.**, **Stutsman, R. ve Atallah, M.** (2005). Translation-based steganography, *International Workshop on Information Hiding*, Springer, s.219–233.

[14] **Topkara, U.**, **Topkara, M. ve Atallah, M.J.** (2006). The hiding virtues of ambiguity: quantifiably resilient watermarking of natural language text through synonym substitutions, *Proceedings of the 8th workshop on Multimedia and security*, ACM, s.164–174.

[15] **Topkara, M.**, **Topkara, U. ve Atallah, M.J.** (2006). Words are not enough: sentence level natural language watermarking, *Proceedings of the 4th ACM international workshop on Contents protection and security*, ACM, s.37–46.

[16] **Topkara, M.**, **Riccardi, G.**, **Hakkani-Tür, D. ve Atallah, M.J.** (2006). Natural language watermarking: Challenges in building a practical system, *Electronic imaging 2006*, International Society for Optics and Photonics, s.60720A–60720A.

[17] **Stutsman, R.**, **Grothoff, C.**, **Atallah, M. ve Grothoff, K.** (2006). Lost in just the translation, *Proceedings of the 2006 ACM symposium on Applied computing*, ACM, s.338–345.

[18] **Taskiran, C.M.**, **Topkara, U.**, **Topkara, M. ve Delp, E.J.** (2006). Attacks on lexical natural language steganography systems, *Electronic Imaging 2006*, International Society for Optics and Photonics, s.607209–607209.

[19] **Bolshakov, I.A.** (2004). A method of linguistic steganography based on collocationally-verified synonymy, *International Workshop on Information Hiding*, Springer, s.180–191.

[20] **Bergmair, R.** (2004). Towards linguistic steganography: A systematic investigation of approaches, systems, and issues, *Final year thesis, B. Sc.(Hons.) in Computer Studies, The University of Derby*.

[21] **Bergmair, R.** (2007). A comprehensive bibliography of linguistic steganography, *Electronic Imaging 2007*, International Society for Optics and Photonics, s.65050W–65050W.

[22] **Bergmair, R. ve Katzenbeisser, S.** (2006). Content-aware steganography: about lazy prisoners and narrow-minded wardens, *International Workshop on Information Hiding*, Springer, s.109–123.

[23] **Meral, H.M.**, **Sankur, B. ve Ozsoy, A.S.**, (2006). Watermarking tools for Turkish texts, 2006 IEEE 14th Signal Processing and Communications Applications.

[24] **Meral, H.M.**, **Sevinc, E.**, **Ünkar, E.**, **Sankur, B.**, **Özsoy, A.S. ve Güngör, T.** (2007). Syntactic tools for text watermarking, *Electronic Imaging 2007*, International Society for Optics and Photonics, s.65050X–65050X.

[25] **Meral, H.M.**, **Sankur, B.**, **Özsoy, A.S.**, **Güngör, T. ve Sevinç, E.** (2009). Natural language watermarking via morphosyntactic alterations, *Computer Speech & Language*, *23*(1), 107–125.

[26] **Zielińska, E.**, **Mazurczyk, W. ve Szczypiorski, K.** (2014). Trends in steganography, *Communications of the ACM*, *57*(3), 86–95.

[27] **Katzenbeisser, S. ve Petitcolas, F.** (2000). *Information hiding techniques for steganography and digital watermarking*, Artech House, Inc.

[28] **Wilson, A.**, **Blunsom, P. ve Ker, A.D.** (2014). Linguistic steganography on twitter: hierarchical language modeling with manual interaction, *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, s.902803–902803.

[29] **Oflazer, K. ve Kuruöz, İ.** (1994). Tagging and morphological disambiguation of Turkish text, *Proceedings of the fourth conference on Applied natural language processing*, Association for Computational Linguistics, s.144–149.

[30] **Yuret, D. ve Türe, F.** (2006). Learning morphological disambiguation rules for Turkish, *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, Association for Computational Linguistics, s.328–334.

[31] **Bilgin, O.**, **Çetinoğlu, Ö. ve Oflazer, K.** (2004). Building a wordnet for Turkish, *Romanian Journal of Information Science and Technology*, *7*(1-2), 163–172.

[32] **Kosowski, A. ve Manuszewski, K.** (2004). Classical coloring of graphs, *Contemporary Mathematics*, *352*, 1–20.

[33] **Stolcke, A.** *ve diğerleri* (2002). SRILM-an extensible language modeling toolkit., *Interspeech*, cilt2002, s.2002.

[34] **Sak, H.**, **Güngör, T. ve Saraçlar, M.**, (2008). Turkish language resources: Morphological parser, morphological disambiguator and web corpus, Advances in natural language processing, Springer, s.417–427.

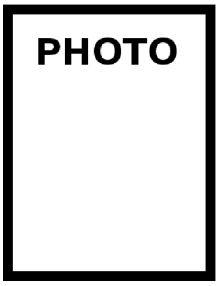
[35] **Kuznetsov, Y.A.** (2013). *Elements of applied bifurcation theory*, cilt112, Springer Science & Business Media.

[36] **Stojanovski, T. ve Kocarev, L.** (2001). Chaos-based random number generators-part I: analysis [cryptography], *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *48*(3), 281–288.

[37] **Kerckhoffs, A.** (1883). La cryptographie militaire, *Journal des sciences militaires*, 5–83.

[38] **Chang, C.Y. ve Clark, S.** (2010). Practical linguistic steganography using contextual synonym substitution and vertex colour coding, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, s.1194–1203.

ÖZGEÇMİŞ



Ad Soyad: Osman Boyacı

Doğum Tarihi ve Yeri: 04.08.1989 Kütahya

E-Posta: boyacios@itu.edu.tr

ÖĞRENİM DURUMU:

- **Lisans:** 2013, İstanbul Teknik Üniversitesi, Elektrik-Elektronik Fakültesi, Elektronik Mühendisliği
- **Lisans:** 2013, İstanbul Teknik Üniversitesi, Bilgisayar ve Bilişim Fakültesi, Bilgisayar Mühendisliği

MESLEKİ DENEYİMLER VE ÖDÜLLER:

- 2013-2014 yılları arası SIEMENS'te gömülü yazılım mühendisi olarak çalıştı.
- 2014-2015 yılları arası TÜBİTAK UEKAE'de araştırmacı olarak çalıştı.
- 2015'ten bu yana ERARGE'de arge mühendisi olarak çalışmaktadır.

YÜKSEK LİSANS TEZİNDEN TÜRETİLEN YAYINLAR:

- Boyacı O., Tantuğ A. C. 2017. A Random Number Generation Method Based on Discrete Time Chaotic Maps. *IEEE International Midwest Symposium on Circuits and Systems*, August 6-9, 2017 Boston, MA, USA.