Avrupa Bilim ve Teknoloji Dergisi
Sayı 18, S. 656-663, Mart-Nisan 2020

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 18, pp. 656-663, March-April 2020

Research Article

Siberbiyogüvenlik Uygulamalarında DNA Dizilimleri için Özet Algoritmaları Karşılaştırılması

Esma Ergüner Özkoç[1*], Mike Mannion[2]

[1] Başkent Üniversitesi, Ticari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Ankara, Türkiye (ORCID: 0000-0003-1728-5930)
[2] Glasgow Caledonian Üniversitesi, School of Engineering & Built Environment, Glasgow, UK (ORCID: 0000-0003-2589-3901)

(İlk Geliş Tarihi 4 Kasım 2019 ve Kabul Tarihi 19 Mart 2020)

(**DOI:** 10.31590/ejosat.642275)

ATIF/REFERENCE: Özkoç, E. E., & Mannion, M. (2020). Siberbiyogüvenlik Uygulamalarında DNA Dizilimleri için Özet Algoritmaları Karşılaştırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (18), 656-663.

Öz

DNA'nın biyolojik bir hesaplama, depolama malzemesi olarak kullanılmasına yönelik ilgi gün geçtikçe artmaktadır. DNA'nın bu şekilde kullanımı ile bilgiler DNA kodları olarak kodlanabilir, iletilebilir ve saklanabilir. DNA'nın, depolama yoğunluğu kabiliyeti, maliyet ve bilgiye erişim gibi kriterlerle değerlendirildiğinde, silikon bazlı yaklaşımlardan daha etkili ve verimli olduğu düşünülmektedir. Bununla birlikte, DNA dizilerini kullanarak bilgi temsili, kodlanmış bilginin saldırıya uğramasını önlemek için yeni teknik güvenlik zorlukları getirmektedir.

Bu çalışmada, farklı sabit uzunluklardaki DNA dizilerinin bütünlüğünü sağlamak için, yaygın olarak kullanılan dört metin tabanlı özet algoritmasının (MD5, SHA1, SHA2-256 ve SHA2-512) uygunluğu araştırılmıştır. Değerlendirme kriteri olarak açık metin duyarlılığı ve çalışma zamanı kullanılmıştır. Sonuçlar, her bir algoritmanın güçlü ve zayıf yanlarının olduğunu ancak genel olarak SHA2-512 algoritmasının açık metin duyarlılığında ve MD5 algoritmasının derleme zamanında daha iyi performansa sahip olduğunu göstermektedir.

Anahtar Kelimeler: Siberbiyogüvenlik, Biyogüvenlik, Sibergüvenlik, Hash algoritmaları, DNA dizisi.

Comparison of Hash Algorithms for DNA Sequences for Cyberbiosecurity Applications

Abstract

There is increasing interest in using DNA as a biological computation storage material in which information will be encoded, transmitted and stored as DNA sequences. DNA is perceived to be more effective and efficient than silicon-based approaches when evaluated using criteria such as storage density capability, cost and access to information without using special equipment and resilience to material change. However, knowledge representation using DNA sequences brings new technical security challenges for preventing the encoded knowledge from being hacked.

In this paper, we examine the suitability of four commonly used text based hash algorithms (MD5, SHA1, SHA2-256 and SHA2-512) to evaluate the integrity of DNA sequences of different fixed lengths. The criteria used for evaluation were plain text sensitivity and compile time. Our results show that each algorithm has strengths and weaknesses but in general the SHA2-512 algorithm performs better on plain text sensitivity and the MD5 algorithm performs better on compile time.

Keywords: Cyberbiosecurity, Biosecurity, Cyber security, Hash algorithms, DNA sequnece.

[*] Esma Ergüner Özkoç: Başkent Üniversitesi, Ticari Bilimler Fakültesi, Yönetim Bilişim Ssitemleri Bölümü, Ankara, Türkiye, ORCID: 0000-0003-1728-5930, eozkoc@baskent.edu.tr

1. Giriş

Yaşam bilimlerindeki güvenlik politikaları iki farklı kategoride incelenebilir: (i) Biyokoruyucu (biosafety) ve (ii) biyogüvenlik (biosecurity) [1]. Biyolojik ajanların laboratuvarlardan çevreye yanlışlıkla salınmasını veya patojenlere farkında olmadan maruz kalmayı önlemek için biyokoruyucu politikaları oluşturulmuştur. Biyokoruyucu önlemlerine örnek olarak koruyucu kıyafetler, sterilizasyon işlemleri verilebilir. Biyogüvenlik ise, bulaşıcı hastalık ajanlarının veya bunların yaratılması, üretimi ve kasıtlı veya kazara salınması yoluyla insanlara, hayvanlara, bitkilere ve çevreye zarar verebilecek olan bilimin kötüye kullanımı ile ilgili riskleri azaltmaya yöneliktir [2]. Türkiyede, 5977 sayılı Biyogüvenlik Kanunu ise "Bilimsel ve teknolojik gelişmeler çerçevesinde, modern biyoteknoloji kullanılarak elde edilen genetik yapısı değiştirilmiş organizmalar ve ürünlerden kaynaklanabilecek riskleri engellemek, insan, hayvan ve bitki sağlığı ile çevrenin ve ekolojik çeşitliliğin korunması, sürdürülebilirliğinin sağlanması amacıyla biyogüvenlik sisteminin kurulması ve uygulanması, bu faaliyetlerin denetlenmesi, düzenlenmesi ve izlenmesi ile ilgili usul ve esasları belirlemek" olarak 26.09.2010 tarihinde yürürlüğe girmiştir [3]. Kısaca, biyokoruyucu ve biyogüvenlik politikaları, sınırlı sayıda biyolojik tehdidi önlemek için tasarlanmıştır. Oysaki halihazırda; kavramsal olarak düzenlenmiş patojenlerin genomik dizilerinden oluşturulan biyolojik silahların geliştirilmesi (Gen sentezi teknolojileri), dizildiğinde bir bilgisayarı uzaktan kontrol edebilecek veri dosyası oluşturan bir DNA örneği tasarlanması, biyoinformatik kaynaklarında mevcut olan DNA dizilerinin doğada bulunmayan biyolojik tehditler oluşturmak için kullanılması [4] ve bu saldırıları geliştirmek için yazılım araçlarının kullanılması gibi tehditler mevcuttur. Başka bir ifade ile günümüzde, biyoteknoloji endüstrisinde, ürün geliştirme iş akışlarının hesaplamalı ve deneysel boyutları arasındaki karmaşık ilişkilerden doğan tehditlere karşı korumayı içeren gelişmiş bir güvenlik kültürü yerleşmemiştir. [5].

Biyokoruyucu ve biogüvenlik önlemlerinin ötesinde siber güvenlik ise öncelikle kişisel bilgisayarlardan ve iletişim cihazlarından büyük altyapılara ve bilgisayar ağlarına kadar bilgi teknolojisi tabanlı sistemlerin güvenliğine odaklanan ayrı bir disiplindir. Dünyada geçtiğimiz birkaç yıla kadar, "biyogüvenlik" ile "siber" örtüşmeler ve aralarındaki önemli ilişki çok anlaşılmamış veya çok önemsenmemiştir. Türkiye'de ise bilindiği kadarı ile henüz bu konuda bir çalışma başlatılmamıştır. Biyoteknoloji endüstrisi, siber uzayda veya biyolojik alanda DNA dizilerinin değiştirilmesi ile ilişkili doğal risklere ek olarak, gittikçe artan bir şekilde siber saldırılara karşı zayıflığı olan bilgisayar kontrollü araçlara dayanmaktadır. Bu ilişki yepyeni bir risk kategorisi oluşturmaktadır. Siberbiyogüvenlik, siber uzay ve biyoloji arasındaki sınırda ortaya çıkan yeni riskleri anlamayı amaçlamaktadır. Bu kapsamda Siberbiyovüvenlik, siber güvenlik, siber fiziksel güvenlik ve biyogüvenlik arabiriminde ortaya çıkmış melezleşmiş bir disiplindir [2]. Siberbiyogüvenlik terimini ilk 2018 yılında öneren Peccoud [1] olmasına karşın ilk ayrıntılı tanımını Murch ve ark. [2] "Gelişmekte olan yaşam ve tıp bilimleri siber, siberfiziksel, tedarik zinciri ve altyapı sistemleri arasında veya arayüzlerinde meydana gelebilecek istenmeyen gözetim, ihlal, kötü niyetli ve zararlı faaliyetlere karşı zafiyetlerin anlaşılması ve güvenlik rekabet edebilirlik ve elastikiyet gibi özelliklere karşı tehditlerin önlenmesi hafifletilmesi araştırılması ve ilişkilendirilmesidir" şeklinde yapmıştır. Bu tanım için başlangıç tanımı ifadesini kullanmışlar ve geliştirilmesi gerektiğini vurgulamışlardır. Ayrıca, çeşitliliği ve kapsamı nedeniyle, siberbiyogüvenliğin kendi sistematiğinin oluşturulması gerekliliğini, böylece daha iyi iletişim kurulabileceği, organize edilebileceği ve geliştirilebileceğini belirtmişlerdir. Richardson [6], Murch ve akr.'nın tanımını siber güvenlik ve biyogüvenlik kapsamından farklılaştırmak adına genişletmeyi önermiştir; "Siberbiyogüvenlik, yaşam bilimleri ve dijital dünya ara yüzündeki değerli bilgilerin, süreçlerin ve materyallerin kötü niyetli imhası, kötüye kullanılması veya sömürülmesi potansiyelini genel olarak kötü niyetli teknoloji kullanımı tehdidi bağlamında anlaşılmasını gerektirir."

Dünya, teknolojinin gelişme hızına göre yavaş kalsa da, bir dizi güvenlik önlemi almaya ve sorunları konuşmaya başlamıştır. Örneğin ABD de gen sentezi hizmet sağlayıcıları için tarama kılavuzları [7] geliştirilmekte ve genom düzenleme teknolojilerinin kötü kullanımı hakkında önlemler alınması gerektiği açıklanmıştır.

Yaşam Bilimlerinde yapılan çalışmalarda iyi niyetli olarak herhangi bir önlem almaksızın bilim insanlarının veri ve örnekleri paylaşması çok sık görülen bir durumdur. Fakat bu durum bazı güvenlik risklerini de beraberinde getirmektedir. Bunun yanı sıra DNA dizilerinin günümüzde, şifreleme algoritmalarında anahtar, kimlik doğrulama şemalarında OTP (One time pad) olarak kullanımı gibi kriptografik amaçlarla da kullanımı artmaktadır. Güvenli veri paylaşımı için kriptografik yöntemler mevcuttur. Paylaşılan verinin bütünlüğünün sağlanması, paylaşılan verilerin karşı tarafa eksiksiz veya değiştirilmeden gittiğinden emin olunması için en temel kriptografik yöntem verinin özet değerinin (hash value) hesaplanmasıdır. Bu sebeple bu çalışmada DNA dizilimleri için açık metin duyarlılığı ve çalışma zamanı açısından en uygun metin tabanlı özet fonksiyonunun belirlenmesi amacı ile yaygın kullanılan özet fonksiyonları karşılaştırılması yapılmıştır. DNA dizilimlerini normal bir metinden ayıran durum ise sadece dört harften oluşmasıdır.

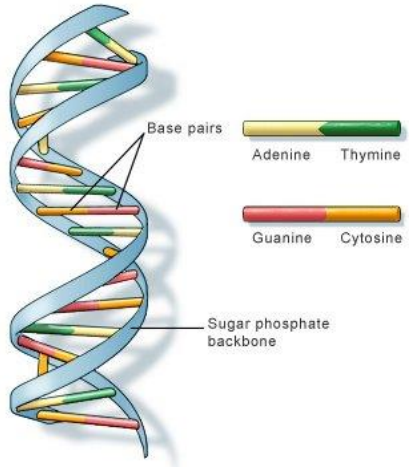
Çalışmanın ilerleyen bölümlerinde öncelikle DNA yapısı ve kullanılan Özet fonksiyonları hakkında kısa bilgi verilecektir. Sonraki bölümlerde ise araştırmada izlenen metottan bahsedilecek ve analiz sonuçları ile beraber tartışma bölümü ile çalışma sonlandırılacaktır.

2. Materyal ve Metot

2.1. DNA Yapısı

DNA (Deoxyribo Nucleic Acid), tüm organizmaların ve bazı virüslerin hayati fonksiyonları ve biyolojik evrimi için gerekli olan genetik bilgiyi taşıyan bir nükleik asittir. Bir DNA'nın tek iplikçiği (ssDNA), nükleobaz olarak adlandırılan bir molekül dizisi şeklindedir. Dört farklı DNA nükleobazı vardır: Adenin (A), Timin (T), Guanin (G) ve Sitozin (C). Bu bazlar DNA alfabesini oluşturur ve tamamlayıcı çiftler halinde gruplandırılır (A-T, G-C) (Şekil 1- Watson-Crick İkili Sarmal Yapısı [8]. DNA iplikçik dizisinin en temel özelliklerinden biri, sırasının olmasıdır: örneğin ATTCA, ACTCA 'ya eşit değildir. Alfabedeki kelimelerin kelimeleri oluşturmak üzere düzenlenmesi gibi, DNA'yı oluşturan nükleobazlar da tüm molekülün yeni bir kopyasını çıkarmak için gerekli tüm bilgileri sağlayan özel bir düzendedir. Bu özel nükleobaz sırası "genetik kod" oluşturur. Her canlının hücrelerinin içine yerleştirilmiş genetik kod, o

canlının "genom" unu oluşturur. DNA içindeki bilgi, hangi proteinden ne kadar üretilmesi gerektiğini belirler. Aynı yapıdaki bir genetik kod ile bir bakteri, bir bitki, bir hayvan veya bir insan da inşa edilebilir. Bu canlıların türünde farklılık gösteren durum, nükleobazların dizilimleri ve sayısıdır. Örneğin bu zamana kadar bilinen en uzun dizilimli hayvan genomu olan Ambystoma mexicanum (Axolotl)'da 32 milyar nükleobaz çifti [9] bulunurken insanda sadece 3.2 milyar nükleobaz çifti [10] bulunmaktadır.



Şekil1 Watson-Crick İkili Sarmal Yapısı

2.2. Kriptografik Özet Fonksiyonları

Verilen farklı uzunluklardaki veriyi belirli uzunlukta bir bit dizisine, özet değerine, dönüştüren fonksiyonlardır. Özet değeri, girdi verinin her bir bitine bağlı olmalıdır. Girdideki bir bitlik değişim bile özet değerini değiştirebilmelidir. Kriptografik bir özet fonksiyonunun birkaç güvenlik gereksinimini sağlaması gerekir; (i) Öngörüntü direnci: Bir özete karşılık gelecek girdi mesajı oluşturulmamalı. (ii) İkinci Öngörüntü direnci: Verilen bir girdi mesajı le özeti aynı olan faklı bir girdi mesajı bulunmaması. (iii) Çakışma Direnci: Aynı özete sahip iki farklı girdi mesajı bulunmaması. Veri bütünlüğünün sağlanması, sayısal imza protokolleri, kimlik doğrulama protokolleri, anahtar türetme gibi farklı güvenlik uygulamalarında kullanımları yaygındır.

En sık kullanılan ve en bilindik iki temel Özet fonksiyonu MD5 ve SHA dır. Tablo1 de bu çalışma dahilinde kullanılan özet algoritmalarının özellikleri verilmiştir.

MD5 (Message Digest Algorithm), MD4'ün güvenlik seviyesinin yeterli olmadığı düşüncesi ile R. Rivest, 1991'de MD4 algoritmasının zayıflıklarının giderilmiş ve güçlendirilmiş bir versiyonu olan MD5'ı önermiştir [11][12]. MD5 algoritmasında girdi mesaj 512 bitlik bloklara ayrılmakta (eğer ayrılmıyorsa mesaja ekleme -padding- yöntemi uygulanır) ve her bir blok kendi içerisinde 16x32 bitlik alt bloklar halinde işlemden geçirilmektedir. Yapılan çalışmalarda MD5 algoritmasının çakışma saldırılarına karşı zayıflığı olduğu ortaya konmuştur [13][14]. Internet standardı RFC1321 de tanımlı olan MD5 2019 itibariyle, kanıtlanmış zayıflıklarına ve güvenlik uzmanlarının karşı çıkmalarına rağmen, yaygın olarak kullanılmaya devam etmektedir.

SHA (Secure Hash Algorithm), SHA özetleme fonksiyonları, NSA (United States National Security Agency) tarafından geliştirilmiş ve NIST (National Institute of Standards and Technology, ABD) tarafından Birleşik Devletler Federal Bilgi İşleme Standardı olarak yayınlanmış bir kriptografik özetleme fonksiyon ailesidir; SHA-0, SHA-1, SHA-2, SHA-3. SHA-2'nin ayrıca özet değer uzunluğu değişen ve aralarında ufak farklar içeren 224, 256, 384 ve 512 bitlik versiyonları mevcuttur. Şubat 2017'de, CWI Amsterdam ve Google, aynı SHA-1 özet değerini üreten iki farklı PDF dosyasını yayınlayarak SHA-1'e çakışma saldırısı yaptıklarını açıklamışlardır. [15] SHA-1 algoritmasının yüksek kırılma olasılığı ile SHA-2 algoritmaları geliştirilmiştir. SHA-2 algoritmaları 2001 yılında FIPS PUB 180-2 standartlarına uygun olarak tasarlanmışlardır. SHA-0 ve SHA-1 için geliştirilmiş ataklara karşı SHA-2 versiyonları için herhangi bir zayıflık belirtmemiştir.

Tablo 1. Özet Algoritmaları Özellikleri

Algoritma	Özet Değer Uzunluğu (bits)	Blok boyutu (bits)	Tur (Round) sayısı	Kullanılan Operasyonlar	Yayın Tarihi
MD5	128	512	64	And, Xor, Rot, Add (mod 2^{32}), Or	1991
SHA-1	160	512	80	And, Xor, Rot, Add (mod 2^{32}), Or	1995
SHA2-256	256	512	64	And, Xor, Rot, Add (mod 2^{32}), Or, Shr	2001
SHA2-512	512	1024	80	And, Xor, Rot, Add (mod 2^{64}), Or, Shr	2001

2.3 Metot

Bu çalışmada Java (security.MessageDigest) kütüphanesinde tanımlı fonksiyon olan MD5, SHA1,SHA2-256, SHA2-512 algoritmaları çeşitli uzunluktaki DNA dizileri üzerinde test edilmiştir. Uygulama, Intel Core™ i7-6700HQ CPU @ 2.6 GHz işlemci ve 16GB RAM özellikli dizüstü bilgisayarda NetBeans bütünleşik geliştirme ortamında gerçekleştirilmiştir.

Sonuçlar ise özet değerinin açık metin duyarlılığına (1) göre değerlendirilmiştir. Özet fonksiyonlarında farklı mesajlar için farklı özet değerlerinin elde edilmesi özet fonksiyonunun açık metin duyarlılığı olarak adlandırılmaktadır. Başka bir ifade ile girdi metindeki küçük değişimler özet değerinde büyük değişimlere neden olmalıdır (Avalance effect).

Açık Metin Duyarlılığı: $\left(\frac{fark(H_0,H_i)}{S_b}\right) * 100$ (1)

H_0 : Orijinal verinin Özet değeri
H_i : i bitlik değişiklik yapılmış olan verinin Özet değeri
S_b : Özet değeri uzunluğu
Örneğin, rasgele üretilmiş 256 bp (baz çifti) DNA dizisi "M" için;
M:"TCTCTCCCCGCGATGAGAATTTCCTACAGTTGGAAGCTGCTGGCAGAGGGGGTCCACGGCGGCTTTGTAAACGTGGC
TACGTGGCTAATAACACCTAAATTGGTACTCCTGGCGTAAGCAATATGGAGGATGTCCCGCTAGTCGTAGAGCGGTCTAC
ATAGTGAACAAGGTACCTTCAGTACGTCTGGGACTTTGGCTCAATAACTGGCTCAATATGCGATGGGCACTGCGCGCCAG
GCAGTGCTGCGAACTGATA"

H_0: ab9ddaf5bab8aa2190e8f6b3768d9ae0f3b7dd9c
H_2: b5a46bd5f250a85125425b7776a669652edce7c9
S_b:40
Olduğu durumdaAçık Metin Duyarlılığı: 87,5 olarak hesaplanmaktadır.

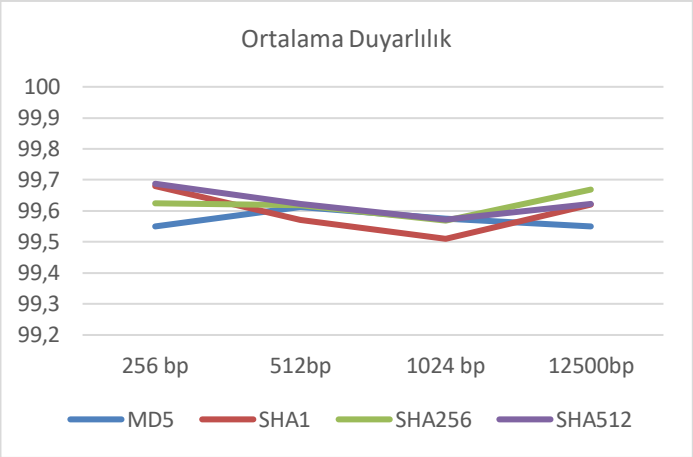
3. Analiz Sonuçları ve Tartışma

Özet fonksiyonlarını karşılaştırmak için, rasgele üretilmiş 256, 512, 1024, 12500 bp uzunluğunda farklı DNA dizileri girdi veri olarak özet fonksiyonuna tabi tutulmuştur. Farklı girdi uzunlukları seçilmesinin sebebi DNA dizilerinin farklı amaçlar için kullanımının yaygınlaşmasıdır. Örneğin DNA dizilimleri bir doğrulama şemasında Tek-Kullanımlık-Şerit (OTP, One time pad) olarak kullanıldığı gibi şifreleme algoritmalarında anahtar olarak da kullanılmaktadır [16].Bunun yanı sıra insan genomunda bulunan genlerin ortalama uzunluğu 10.000 bp ve 15.000 pb arasında değişmektedir [17]. Araştırmacıların herhangi bir genin baz diziliminin bütünlüğünü sağlayabilecekleri bir siberbiyogüvenlik uygulamasında ihtiyaç duyulabilecek en uygun özet fonksiyonunun belirlenmesi amacı ile de 12500 bp uzunluklu DNA dizilimi seçilmiştir.

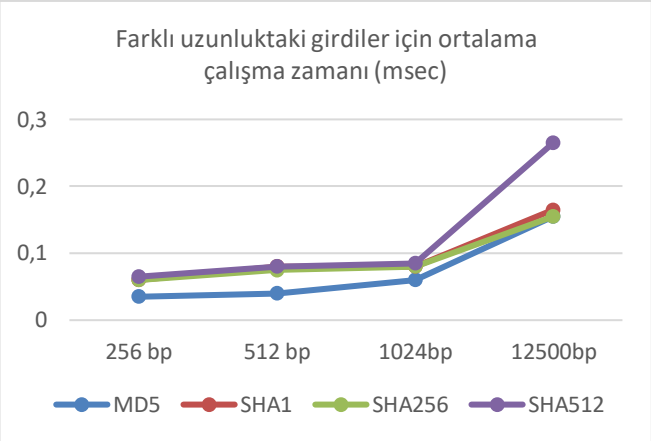
Çalışmada her bir girdi uzunluğu için 10'ar farklı DNA dizisi rasgele üretilmiş (sadece büyük harf) olup sonuçlar bu girdilere karşı alınan özet değerlerinin açık metin duyarlılığının ortalaması olarak hesaplanmıştır (Şekil 4). Her bir DNA dizisi için ilk 50 bp (most significant bit) de değişiklik yapılmıştır. Bunun için girdi DNA diziliminde bulunan A bazının T ile, T bazının G ile, C bazının A ile, ve son olarak G bazının C ile yerleri değiştirilmiştir.

Alınan sonuçlara göre algoritmaların, farklı uzunluktaki DNA dizileri için ortalama duyarlılığı Şekil 2 de verilmiştir.

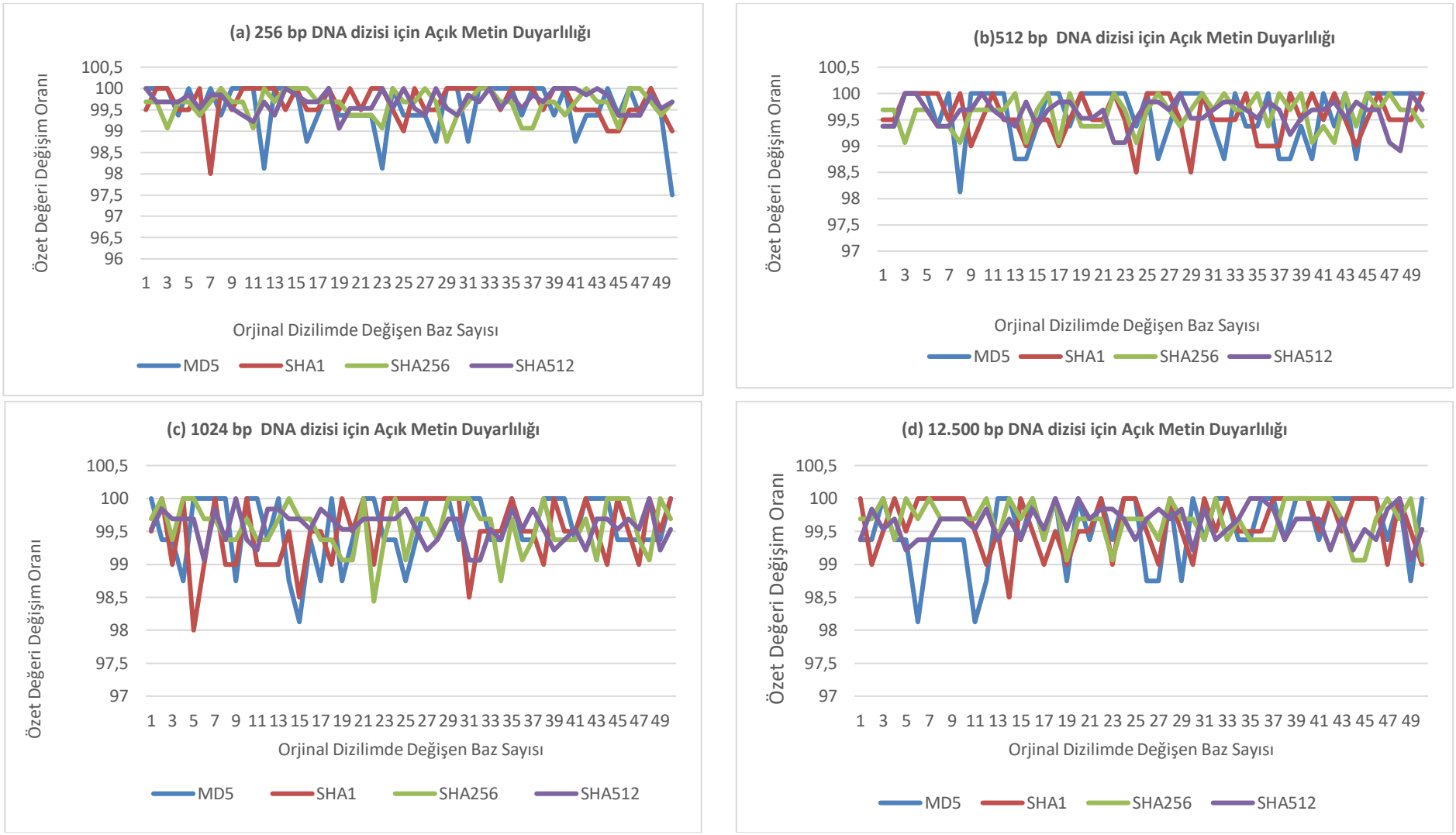
Algoritmalar farklı uzunluktaki girdiler ile 200'er kere çalıştırılmış ve ortalama çalışma zamanı hesaplanmış ve ortalama çalışma zamanları Şekil 3 de karşılaştırmalı olarak verilmiştir.



Şekil 2 Farklı uzunlukta DNA dizileri için özet algoritmalarının ortalama açık metin duyarlılıkları



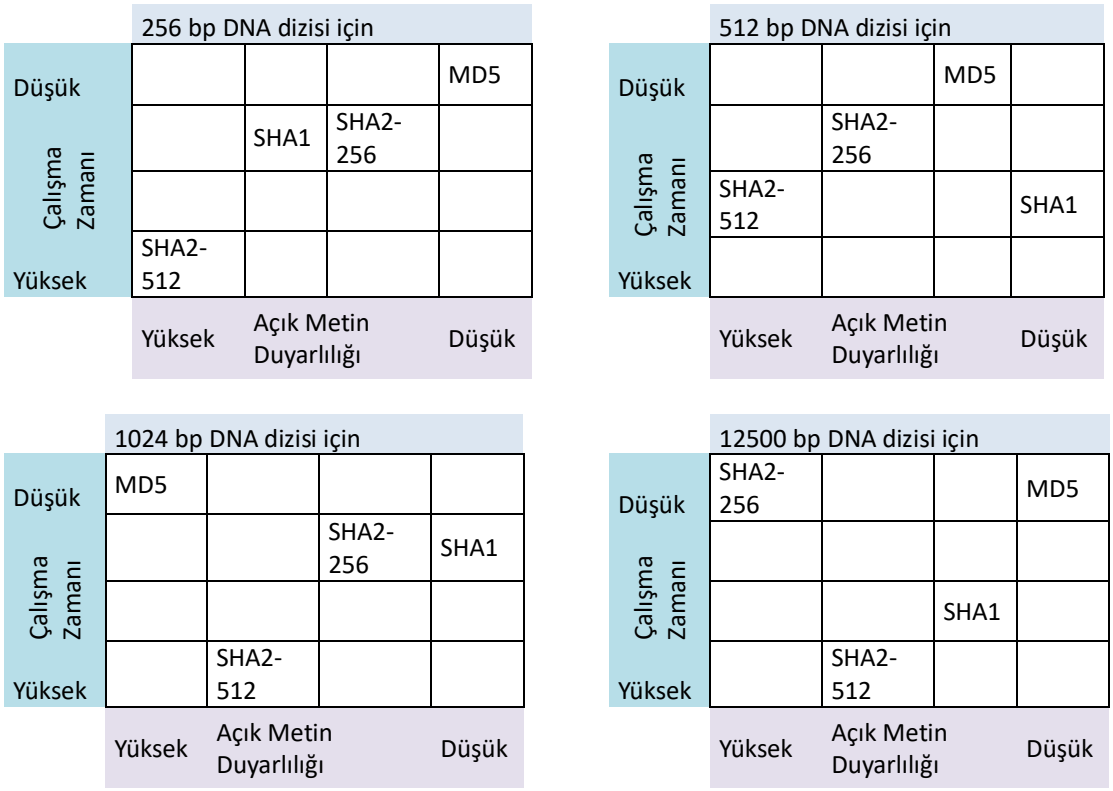
Şekil 3 Farklı uzunlukta DNA dizileri için özet algoritmalarının ortalama çalışma zamanı (200 çalışma için)



Şekil 4- Farklı DNA dizisi uzunlukları (a:256 bp,b:512 bp,c:1024 bp,d:12500 bp) için 50 bp'e kadar her bir algoritmanın hesaplanan açık metin duyarlılığı grafikleri

Siberbiyogüvenlik kapsamında herhangi bir DNA diziliminin bütünlüğünün sağlanması ciddi önem taşımaktadır öyle ki tek bir bazın değişmesi sonuçları katastrofik olabilir. Bu sebeple bu çalışmada, günümüzde birçok güvenlik uygulaması için yaygın olarak kullanılan Özet algoritmalarının, sadece 4 harften oluşan bir alfabe ile yazılan metinler yani DNA dizilimleri için en uygun olanı belirlenmeye çalışılmıştır. Algoritmaların yaygın kullanımlarına rağmen güvenlik açıkları olduğu tespit edilmiştir fakat kolay erişilebilinirlikleri ve henüz DNA dizilimleri için özel tasarlanmış özet algoritmalarının var olmaması sebebi ile içlerinden en duyarlı olan araştırılmıştır.

Analiz sonuçlarına göre Şekil 5 de görüldüğü gibi farklı uzunluktaki DNA dizileri için en kısa çalışma zamanı MD5 algoritmasına aittir. MD5 algoritmasının çalışma zamanı olarak en yakın rakibi ile arasındaki fark ortalama 0.02 milisaniyedir. 256bp ve 512 bp uzunluklarındaki DNA dizileri için açık metin duyarlılığı en yüksek algoritma SHA2-512 iken 1024pb için MD5, SHA2-512 den sadece 0.003 fark ile daha yüksek duyarlılığa sahiptir. 12500 bp uzunluklu DNA dizisi için en yüksek açık metin duyarlılığı SHA2-256 algoritmasına aittir. Ayrıca SHA2-256 bu uzunluktaki bir DNA dizilimi için MD5 algoritması ile aynı sürede özet değeri hesaplamıştır. Her ne kadar MD5 algoritması çalışma zamanı açısından üstünlüğe sahip olsa da bilinen güvenlik zafiyetlerinden dolayı tercih edilmemesi önerilmektedir. Yine SHA1'in DNA dizisi uzunluğu arttıkça açık metin duyarlılığının düştüğü görülmekte ve çakışma saldırısı zayıflığı mevcut olduğu için uzun DNA dizilerinde kullanımı önerilmemektedir. Genelleyecek olursak SHA2-512 algoritması test edilen tüm DNA dizisi uzunlukları için makul bir açık metin duyarlılığına sahiptir. Yukarıda verilen çalışma bulgularına göre, özet algoritması seçimi, uygulamanın hız ve güvenlik gereksinimlerine ve kullanılan DNA dizisinin uzunluğuna göre yapılmalıdır sonucu çıkmaktadır.



Şekil 5- Algoritmaların farklı uzunluktaki DNA dizileri için Çalışma zamanı ve Açık metin Duyarlılıkları karşılaştırması

4. Sonuç

Siberbiyogüvenlik, dünyada henüz birkaç yıllık mazisi olan yeni bir disiplindir. Günümüzde organizmalar için genetik devrelerin ve hücre içermeyen sistemlerin tasarımı, robotiklerin, mikro-akışkanların, sentetik metabolik mühendisliğin, doku mühendisliğinin ortaya çıkması, yeni siber-güvenlik riskleri ve benzersiz tehdit alanları yaratmaktadır[18];[6], [19],[20],[2].

Bu gelişmeler, laboratuvar otomasyon tekniklerinde üretimin arttırılması için geleneksel maliyetin düşürülmesi için yapay zeka, makine öğrenme, robotik alanlarına entegrasyon ile sağlanacaktır. Doğal olarak bu süreçte siberbiyogüvenlik önemli bir gereksinimdir. Siber biyogüvenliğin sağlanması için öncelikle dünyada yeni politikalar ve düzenlemeler yapılmaya başlanmıştır. Teknik açıdan ise siber güvenlik çözümlerinin siberbiyogüvenlik kapsamında adaptasyonu yeni algoritmaların önerilmesi beklenmektedir. Bu çalışmada

transfer edilen verinin bütünlüğünün sağlanması amacıyla kullanılan özet algoritmalarının sadece dört karakterden oluşan DNA dizilimleri üzerindeki duyarlılığı araştırılmıştır. Özet algoritmalarından MD5, SHA1 ve SHA2'nin iki versiyonu (SHA2-256, SHA2-512) yaygın kullanımları ve kolay erişilebilirlikleri açısından seçilmiştir. Yapılan çalışmaya göre, algoritmaların ortalama duyarlılığı % 99 un üzerinde olmasına karşın en yüksek açık metin duyarlılığı farklı uzunluktaki DNA dizilerinde farklılık göstermektedir. SHA2-512 açık metin duyarlılığında test edilen tüm DNA uzunlukları için iyi sonuç vermiştir genellemesi yapılabilir. Bunun yanı sıra SHA2-512nin bilinen bir çakışma zafiyetinin olmaması da algoritmayı bir adım öne çıkarmaktadır. Diğer taraftan MD5 algoritması tüm farklı uzunluktaki DNA dizileri için en kısa zamanda özet değerleri hesaplamıştır. Fakat MD5 algoritmasının bilinen güvenlik zayıflığı mevcuttur. Sonuç olarak seçilecek özet algoritması uygulamanın gereksinimlerine göre farklılık göstermektedir.

Çalışmada, en yaygın kullanılan metin tabanlı özet algoritmaları test edilmiştir. Gelecek çalışmalar kapsamında bir DNA dizisini görüntüye dönüştürerek, en yaygın kullanılan görüntü tabanlı özet yaklaşımları duyarlılık açısından ele alınması planlanmaktadır.

Kaynakça

[1] Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., & Raman, S. (2018). Cyberbiosecurity: from naive trust to risk awareness. Trends in biotechnology, 36(1), 4-7.

[2] Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., & Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. Frontiers in bioengineering and biotechnology, 6, 39.

[3] Anonim (2010): Biyogüvenlik Kanunu. 26 Mart 2010 tarihli Resmi Gazete, Sayı: 27533

[4] Tumpey, T. M., Basler, C. F., Aguilar, P. V., Zeng, H., Solórzano, A., Swayne, D. E., ... & Garcia-Sastre, A. (2005). Characterization of the reconstructed 1918 Spanish influenza pandemic virus. science, 310(5745), 77-80.

[5] Turner, G. (2019, May). The Growing Need for Cyberbiosecurity. In InSITE 2019: Informing Science+ IT Education Conferences: Jerusalem (pp. 207-215).

[6] Richardson, L. C., Connell, N. D., Lewis, S. M., Pauwels, E., & Murch, R. S. (2019). Cyberbiosecurity: a call for cooperation in a new threat landscape. Frontiers in Bioengineering and Biotechnology, 7.

[7] Adam, L., Kozar, M., Letort, G., Mirat, O., Srivastava, A., Stewart, T., ... & Peccoud, J. (2011). Strengths and limitations of the federal guidance on synthetic DNA. Nature biotechnology, 29(3), 208.

[8] U.S.National Library of Medicine, https://ghr.nlm.nih.gov/primer/basics/dna internet adresinden 11.10. 2019 tarihinde edinilmiştir.

[9] Nowoshilow, S., Schloissnig, S., Fei, J. F., Dahl, A., Pang, A. W., Pippel, M., ... & Falcon, F. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature, 554(7690), 50.

[10] National Human Genome Research Institute (NHGRI). https://www.genome.gov/human-genome-project/Completion-FAQ internet adresinden 11.10. 2019 tarihinde edinilmiştir.

[11] Rivest, R.L., (1991) "The MD5 message digest algorithm," Presented at the rump session of Crypto'91.

[12] Rivest, R.L. (1992)"The MD5 message-digest algorithm," Request for Comments (RFC) 1321, Internet Activities Board, Internet Privacy Task Force, April 1992.

[13] [Stevens, M. (2007). On collisions for MD5.

[14] Dougherty, C. R. (2009). Vulnerability Note VU# 836068 MD5 vulnerable to collision attacks. Retrieved August, 26, 2009.

[15] Stevens, M., Bursztein, E., Karpman, P., & Albertini, A. (2017). Announcing the first SHA1 collision (2017). URL: https://security. googleblog. com/2017/02/announcing-first-sha1-collision. html.

[16] Özkoç, E. E. DNA-based user authentication schemes for wireless body area network. e-Society 2018, 217.

[17] Harvard University, The Database of Useful Biological Numbers https://bionumbers.hms.harvard.edu/bionumber.aspx?id=104316&ver=1 internet adresinden 11.10. 2019 tarihinde edinilmiştir.

[18] Nielsen, J., & Keasling, J. D. (2011). Synergies between synthetic biology and metabolic engineering. Nature biotechnology, 29(8), 693.

[19] Rollin, J. A., Tam, T. K., & Zhang, Y. H. P. (2013). New biotechnology paradigm: cell-free biosystems for biomanufacturing. Green chemistry, 15(7), 1708-1719.

[20] Kiss, A. A., Grievink, J., & Rito-Palomares, M. (2015). A systems engineering perspective on process integration in industrial biotechnology. Journal of Chemical Technology & Biotechnology, 90(3), 349-355.