BİTİRME PROJESİ

Haftalık Rapor – 15.10.2021

15 EKİM 2021

KIRIKKALE ÜNİVERSİTESİ – BİLGİSAYAR MÜHENDİSLİĞİ İÖ

AHMET MUNGAN – 160255081

İÇİNDEKİLER

ÖZET

Sesin bir sonraki aşamalarda yapılacağı üzere; matematikselleştirildiği varsayılarak, bazı kriptosistemler ile damgalanacak verinin uygunluk hesaplamaları yapılmıştır. Bu çerçevede lineer denklemleri, lineer bağımlı sistemleri ve lineer bağımsız sistemleri kapsayan bir model ortaya koyulmaya çalışılmıştır. Bu yöntemin tersine mühendislik ile yorumlaması da yapılmıştır.

SESİN KRİPTOSİSTEMLER İLE DAMGALANMASI

Sesin içerisinde veri gizlenmesi veri bilimi ile bilinen yöntemler ile mümkün olmadığını bilmekteyiz. Fakat seste veri gizlemenin bir çok yöntemi mevcut. Geçen hafta raporu itibariyle yayılmış dalgalarla veri gizleme işleminin teknik olarak nasıl yapıldığı söylenmiştir. Bu tarz tekniklerin sayısı çoktur ve araştırılmadan önce, asıl önemli olan kısım bu verinin saklanmasına (şifrelenmesine) değinilecektir.

Veri şifrelemek için aslında kriptoloji biliminin yararlı çıktılarını kullanabilir. Fakat bu çerçevede atlanılmaması gereken bir nokta mevcuttur. Bu şifreleme bir medya ortamı[1] üzerinde gerçekleştirilecektir. Dolayısıyla kriptosistemlerin bir kısmı bu medya ortamında elde edilen ya da işlenen veriler için uygun olmayabilir. Ayrıca bir kriptosistem yöntemi ile verinin gizlenmesi yeterli olmayacaktır. Gizlenmiş verinin mevcut düzende temiz veri midir yoksa farklı bir durum mu söz konusudur incelenecektir. Tüm bu detaylar projenin sonunu ifade ediyor olsa da bu süreçte geriye dönerek çalışmalar da yapılacaktır.

Kriptosistemlerin bilinen başlıca özelliklerinden birisi karakterler üzerinden şifreleme yapmasıdır. Fakat bu karakterleri ASCII kodlamasında bitler haline çevirerek bunların şifrelenmesi de mümkündür. Her karakter bir bit dizisi, her bit ise bir karakterin parçacağı olduğundan; en uygun kriptosistem şudur diye tekelleştirme yapılamaz. Dolayısıyla ses verisi şifrelerken deneme yanılma yöntemiyle en uygun kriptosistem bulunmaya çalışılacaktır.

[1] Buradaki kasıt sestir.

SESİ ŞİFRELEMEK: HILL ŞİFRELEME SİSTEMİ

Gizli anahtarlı[2] bir yöntemdir. Matematikçi Lester S. Hill adını verdiği yöntemde veri bloklara ayrılır. Hill şifreleme sisteminin asıl ve en dikkat çekici kısmı ise gizli anahtarın biçimidir. Gizli anahtarın biçimi; $n * n$ şeklinde bir matristir. Gizli anahtar seçerken, dikkat edilmesi gereken mühim bir husus vardır. Eğer gizli anahtar $mod(26)$'da terslenebilir değilse gizli anahtar seçilemez. Türkçe alfabede $mod(29)$'da terslenebilir olması önemlidir.

Anahtar seçildikten sonra veri n uzunlukta bloklara ayrılır. A anahtarı, P açık verinin bir bloğu ve C şifreli blok olmak üzere:

$$A = \begin{bmatrix} k_{1,1} & \cdots & k_{1,n} \\ \vdots & \ddots & \vdots \\ k_{n,1} & \cdots & k_{n,n} \end{bmatrix}, P_i = \begin{bmatrix} P_1 \\ \vdots \\ P_n \end{bmatrix}, C = \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix}$$

iken, öyle ki;

$$C = \begin{bmatrix} k_{1,1} & \cdots & k_{1,n} \\ \vdots & \ddots & \vdots \\ k_{n,1} & \cdots & k_{n,n} \end{bmatrix} \begin{bmatrix} P_1 \\ \vdots \\ P_n \end{bmatrix} mod(29)$$

ile tanımlanır.

Ayrıca gizli anahtar tersi alınabilir bir matris olduğundan, deşifre algoritması:

$$C = A * P \leftrightarrow A^{-1} * C = A^{-1} * A * P \leftrightarrow A^{-1} * C = I * P \leftrightarrow A^{-1} * C = P$$

şeklinde elde edilir.

Bu çerçevede örnek verilebilmesi açısından gizli bir anahtar seçmek gerekirse ve bu anahtarın terslenebilir olup olmadığı kontrol edilirse, bu anahtara göre şifreleme yapılabilir. Örnek olarak gizli anahtar $A = \begin{bmatrix} 3 & 7 \\ 1 & 2 \end{bmatrix}$ seçilip kontrolü için sırasıyla aşağıdaki lineer bağlantılar kullanılırsa:

$$S_1 \leftarrow 10S_1$$

[2] Gizli anahtar, taraflar arasında anlaşmayla belirlenir. Bu durumda seste şifre/deşifre işlemini yapacak kişi/kişiler kastedilmiştir.

$$S_2 \leftarrow S_2 - S_1$$

$$S_2 \leftarrow 26S_2$$

$$S_1 \leftarrow S_1 - 125S_2$$

bulunacak sonuç: $A^{-1} = \begin{bmatrix} 27 & 7 \\ 1 & 26 \end{bmatrix}$ olarak elde edilir. $mod(26)$'ya göre sadeleştirmesi yapılabilir.

Özet olarak yukarıda hill şifreleme sisteminin matematiksel alt yapısı ve tanımı verilmiştir. Matematiksel kısımlar geçilerek örnek olarak bulunan anahtar kullanılıp "kara kapı" açık metnini ikili bloklar haline getirip şifrelersek:

$$m_1 = "ka", m_2 = "ra", m_3 = "ka", m_4 = "pı"$$

ve

$$C_1 = A * [m_1] = \begin{bmatrix} 3 & 7 \\ 1 & 2 \end{bmatrix} * \begin{bmatrix} 13 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 \\ 13 \end{bmatrix} \rightarrow "ık"$$

$$C_2 = A * [m_2] = \begin{bmatrix} 3 & 7 \\ 1 & 2 \end{bmatrix} * \begin{bmatrix} 20 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 20 \end{bmatrix} \rightarrow "uc"$$

$$C_3 = A * [m_3] = \begin{bmatrix} 3 & 7 \\ 1 & 2 \end{bmatrix} * \begin{bmatrix} 13 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 \\ 13 \end{bmatrix} \rightarrow "ık"$$

$$C_4 = A * [m_4] = \begin{bmatrix} 3 & 7 \\ 1 & 2 \end{bmatrix} * \begin{bmatrix} 19 \\ 10 \end{bmatrix} = \begin{bmatrix} 11 \\ 10 \end{bmatrix} \rightarrow "ıi"$$

ile hesaplanır. Şifreli metin birleştirilirse ve boşluk karakteri aradan alınırsa:

$$C = "ıkucıkıi"$$

şeklinde şifreli metni elde etmiş oluruz.

TERSİNE MÜHENDİSLİK İLE HILL ŞİFRELEME SİSTEMİNİN YORUMLANMASI

Elde edilen metin incelendiğinde, “ıkucıkıı” gibi bir metinden bir çıkarım yapmak zorudur. Zaten hiçbir zaman veri seti bu kadar küçük olamaz. Dolayısıyla veri setini görsel olarak büyütmek[3], tekrara düşerek sadece ufak bir yanılsama ile aşağıdaki adımlar takip edilerek gözlemler yapılabilir. Çok basit bir program koduyla yardım alınıp görsel olarak bir çıktı elde edilebilir.

PyCharm 1

```
C = "ıkucıkıı" * 99
turkSozlugu = {"a":0,"b":0,"c":0,"c":0,"ç":0,"d":0,"e":0,"f":0,
               "g":0,"ğ":0,"h":0,"i":0,"ı":0,"j":0,"k":0,"l":0,
               "m":0,"n":0,"o":0,"ö":0,"p":0,"r":0,"s":0,"ş":0,
               "t":0,"u":0,"ü":0,"v":0,"y":0,"z":0}
for i in C:
    turkSozlugu[i] += 1
print(turkSozlugu)
dosya = open("frekans.txt", "w")
dosya.write(turkSozlugu.__str__())
```

PyCharm 1’de çok basit bir frekans tutucu yapılmıştır. Bu frekans tutucu harflerin ağırlıklarına göre harflerin sayısını arttırmaktadır. Bu algoritmanın çıktısında verilerin özeti alınırsa[4]:

{…'c': 99, 'i': 99, 'ı': 297, 'k': 198, 'u': 99,…}

Şeklinde bir çıktıyla karşı karşıya kalırız. Bu çıktıdan anlaşılacağı üzere 297 ağırlığı ile birinci olan “ı” harfi türkçede frekansı en yüksek “a” harfinin yerine yazılmış olabilir. Elbette bu kanıya ufak bir veri seti ile varmak mümkün değildir. Dolayısıyla bu veri setini büyütmek şifreleme yönteminin tek düze devam etmesi nedeniyle, harf frekansı ile çözülebilecek bir şifre olması sebebiyle, sadece lineer tabanlı değil aynı zamanda akıl yürüterek de bu şifrenin kırılabileceğinden ötürü hill şifreleme sistemi güvenilir değildir. Aynı zamanda kriptosistemlerin bir arada kullanılmasıyla güçlü yöntemler elde edildiği göz önünde bulundurulursa, sadece eldeki veriden bile güçsüz çıkan bir şifreleme tekniğinin kullanılmaması gerekir.

---

[3] Bir veriyi görsel olarak büyütmek onu resmen büyütmek anlamında kullanılmamıştır.
[4] Veri özeti alınmasının sebebi veri setinin küçük ve kendini tekrar etmesindendir.

Bu şifrenin kuvvetli olabileceği gerekçesi ile raporun ikinci kısmında[5] hill yönteminin açıklanması zayıf kalmıştır. Dolayısıyla çarpaz ya da biribiri içerisine geçmiş şifreleme sistemleri blok olarak şifrelenen veriler için daha uygun olacaktır. Aynı zamanda lineer tabanlı şifreleme sistemlerinden bir diğeri playfair şifreleme sistemi zaten türkçe için uygun bir şifreleme değil. Ayrıca hill şifreleme sisteminin harfler üzerinden gidildiği düşünülürse, zaten ses ortamındaki matematiği yakalamak zor olacağı için playfair gibi türkçe karakterlerin tanımı olmayan şifreleme sistemleri çıkmaza sokabilir.

Tüm bu sebep ve çıktıların sonucunda: Hill şifreleme sistemi, ses[6] için lineer çözüm uzayından ötürü uygundur fakat tersine mühendislik teknikleri ile bu şifreleme zayıf kalmaktadır. Dolayısıyla başka bir kriptosistem ile birleştirilerek kullanılması gerekmektedir.

[5] Raporun bahsi geçen kısmına gitmek için tıklayınız.
[6] Sesin matematikselleştirildiği kabul edilmektedir.

ANAHTAR KELİMELER

- ***ASCII*** : Latin alfabesi üzerine kurulu 7 bitlik bir karakter kümesidir.

REFERANS VE KAYNAKÇA

- İTÜBİDB, Şifreleme Yöntemleri – Link için tıklayınız.
- M. Yılmaz, S. Ballı, "Veri Şifreleme Algoritmalarının Kullanımı İçin Akıllı Bir Seçim Sistemi Geliştirilmesi", Uluslararası Bilgi Güvenliği Mühendisliği Dergisi, Cilt: 2, No: 2, Sayfa: 18-28, 2016.
- T. Yerlikaya, E. Buluş, N. Buluş, "Kripto Algoritmalarının Gelişimi ve Önemi", Trakya Üniversitesi, Bilgisayar Mühendisliği.
- W. Stallings, "Kriptografi ve Siber Güvenlik Prensipleri ve Uygulamaları", 4. Baskı, Prentice Hall Publication, 2005
- Kriptolojiye Giriş Ders Notları, Uygulamalı Matematik Enstitüsü, Kriptografi Bölümü, ODTÜ, 2004