VERİ MAHREMİYETİ: SALDIRILAR, KORUNMA VE YENİ BİR ÇÖZÜM ÖNERİSİ

Yılmaz VURAL
Kişisel Verileri Koruma Kurumu, Ankara, Türkiye
yilmazvural@gmail.com

ÖZET

Günlük yaşantımızın ayrılmaz parçası haline gelen elektronik uygulamalar aracılığıyla çeşitliliği ve büyüklüğü her geçen gün artan veriler toplanmakta ve işlenmektedir. Farklı amaçlar için işlenen bu veriler içerisinde kişileri doğrudan veya dolaylı olarak tanımlayan kişisel veriler de yer almaktadır. Kişisel verilerin işlenmesi sırasında gerekli olan idari ve teknik tebirlerin alınmaması veri ihlallerinin yaşanmasına neden olmaktadır. Veri ihlallerinin kişilere, kurumlara ve ülkelere verdiği zararların azaltılması amacıyla mahremiyet koruyucu önlemlerin alınması gerekmektedir. Veri mahremiyeti, veri sahiplerinin mahremiyeti ile veri paylaşımının taraflara sağlayacağı fayda arasındaki en iyi dengeyi bulmaya çalışan zor bir problemdir Bu çalışmada, literatürdeki mahremiyet koruyucu yöntemler incelenmiş, incelenen yöntemlerin güçlü ve zayıf yönleri araştırılmış, veri faydası metrikleri değerlendirilmiş ve mahremiyetle ilgili saldırılar gözden geçirilmiştir. Çalışma kapsamında elde edilen bulgular, yapılan araştırmalar, tespitler ve değerlendirmeler sonucunda veri faydasını gözeterek mahremiyeti sağlamaya yönelik yeni bir veri çoğaltma yaklaşımı sunulmuştur. Önerilen veri çoğaltma yaklaşımının, veri faydasını koruyarak mahremiyet saldırılarını önemli oranda azaltacağı, karşılaşılan olumsuzlukları önleyeceği ve en önemlisi kişisel verilerin korunmasına önemli katkılar sağlayacağı değerlendirilmektedir.

Anahtar Kelimeler: Anonimleştirme; Kişisel verilerin korunması, Mahremiyet koruyucu veri paylaşımı; Mahremiyet modelleri; Mahremiyet saldırıları; Veri çoğaltma.

DATA PRIVACY: ATTACKS, PROTECTION AND NEW APPROACH

ABSTRACT

In this study, the literature on the privacy preserving methods was examined, the strengths and weaknesses of the existing methods were represented, and privacy related attacks were reviewed. This comprehensive analysis on the existing privacy preserving solutions clearly shows that a new complete and efficient approach needs to be explored for peer-to-peer anonymized data sharing. Therefore; a new method has been suggested. The method is based on record population for providing privacy and preserving and purposes a sensitive balance between data benefits and data privacy. The results indicate that populating data sets significantly reduces the possibility of privacy violations.

Keywords: Anonymization; Data privacy; Data population; Personal data protection, Privacy awareness; Privacy preserving data sharing; Privacy models; Privacy attacks.

1. GİRİŞ (INTRODUCTION)

Günlük yaşamın ve iş hayatının ayrılmaz parçası olan elektronik uygulamaların kişisel (sosyal medya, oyunlar, sayısal asistanlar, vb.) ve kurumsal (sağlık, nüfus, finans, eğitim, yerel yönetimler, mülkiyet, vb.) alandaki kullanımları hızla yaygınlaşmaktadır [1]. Kurumsal veya kişisel amaçlara göre belirli bir işin yerine getirilmesiyle bağlantılı olarak farklı gerekliliklere (kanuni, sözleşme, açık rıza, vb.) göre günlük ve iş hayatının düzen içinde sürdürülmesi, hizmetlerin etkin biçimde sunulması, geliştirilmesi, dağıtımı ve

pazarlanması amacıyla veriler çoğunlukla siber uygulamalar aracılığıyla toplanmaktadır.

Toplanan bu verilerin işlenmesi gerçek kişilere, kurumlara ve araştırmacılara önemli fırsatlar sunarken beraberinde yeni tehditleri, tehlikeleri ve sorumlulukları da getirmektedir [2]. Kişileri doğrudan veya dolaylı olarak tanımlayan kişisel verilerin de paylaşılan veriler içerisinde olması mahremiyet koruyucu tedbirlerin alınmasını zorunlu hale getirmektedir. Mahremiyet koruyucu önlemlerin alınmasında sadece verisi işlenen gerçek kişilerin korunması değil aynı zamanda paylaşılacak verinin taraflara sağlayacağı faydanın da dikkate alınması gerekmektedir. Veri mahremiyeti, veri sahiplerinin mahremiyeti ile veri paylaşımının taraflara sağlayacağı fayda arasındaki en iyi dengeyi bulmaya çalışan zor bir problemdir [3].

Mahremiyet bilincinin oluşmaması, önlemlerin yeterli düzeyde alınmaması birçok mahremiyet odaklı saldırıya davet çıkarmaktadır. Mahremiyet saldırıları, kişisel verilerin hukuka aykırı olarak işlenmesiyle gerçek kişilerin ayrımcılığa veya mağduriyete uğramasına, kurum, kuruluş ve ülkelerin zarara uğramasına yol açmaktadır. 1990 yılında ABD'de sayım uygulamasıyla toplanan cinsiyet, posta kutusu ve doğum tarihi gibi doğrudan kişileri tanımlamayan verilerin kullanılarak ABD nüfusunun %87'sinin kimliklerinin tespit edilebileceği Sweeney tarafından raporlanmıştır [4].

ABD'de 37 eyaletteki sağlık bilgilerinin toplanması için görevlendirilen ABD Sağlık Kuruluşları Birliği (NAHDO) 1996 yılında önemli bir mahremiyet saldırısına maruz kalmıştır. Saldırganlar hasta bilgileri ile daha önce yayınlanan seçmen kayıt listeleri arasında bağlantı kurarak gerçek kişilerin kimliklerini ve özel nitelikli kişisel verilerini ifşa etmişlerdir [5,6]. 2006 yılında AOL 20 milyon web sorgusunu içeren 650 bin kullanıcının 3 aylık arama kayıtları anonim kimlik numarasıyla araştırmacılar için yayınlamış ancak yayınlanan kayıtlardan kişilerin kimliklerine ulaşılmıştır [3,7]. Dünyada olduğu gibi ülkemizde de yeterli önlemler alınmaması, mahremiyet bilinci eksikliği ve kişisel verilerin korunmamasına bağlı olarak veri ihlalleri yaşanmaktadır [8,9].

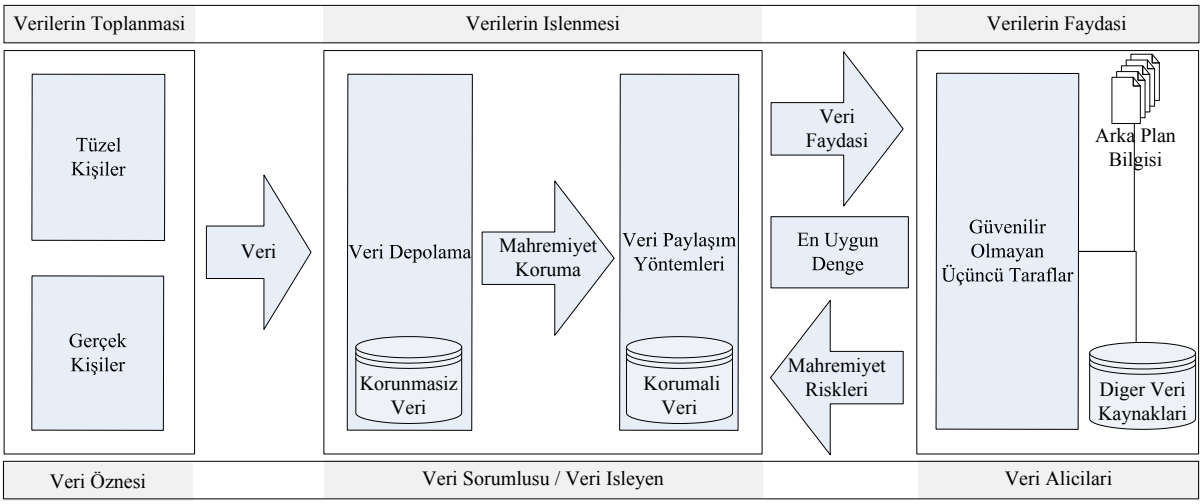
Veri ihlalleri, gerçek kişilerin ayrımcılığa veya mağduriyete uğramasından toplumları ve ülkeleri hedef alan kitlesel tehditlere kadar geniş bir alanda etkisini göstermekte ve sayıları artarak devam etmektedir. Mahremiyet odaklı veri ihlallerinin en aza indirgenmesi amacıyla mevcut saldırıların ve eksikliklerin analiz edilerek veri faydası ihtiyaçlarına göre yeni mahremiyet koruyucu modellere ihtiyaç duyulmaktadır [10]. Veri mahremiyetinin korunmasına yönelik teknik ve idari çalışmaların yapılarak kişisel verilerin hukuka uygun olarak işlenmesindeki uygunsuzlukların giderilmesi kişisel verilerin korunması açısından önemlidir.

Bu çalışmada, Bölüm 2'de veri mahremiyeti konusunda kapsamlı literatür incelemesi yapılarak mevcut yöntemlerin güçlü ve zayıf yönleri ortaya koyulmuş, Bölüm 3'de nitelik ve kimlik ifşasına yönelik mahremiyetle ilgili saldırılar gözden geçirilmiş, Bölüm 4'de veri faydasının ölçülmesine yönelik metrikler incelenmiş, Bölüm 5'de kayıt çoğaltma yöntemiyle veri mahremiyetinin korunmasına yönelik yeni bir yaklaşım sunulmuş, Bölüm 6'da ise bu çalışmanın sonuç ve değerlendirmelerine yer verilerek gelecekteki çalışmalar hakkında bilgi verilmiştir.

2. VERİ MAHREMİYETİ (DATA PRIVACY)

Veri mahremiyeti, veri sorumluları tarafından toplanan verilerin yaşam döngüsü içerisinde hukuka ve etik kurallara uygun olarak işlenmesiyle veri öznelerinin korunması olarak tanımlanabilir. Verilerin toplanması ve paylaşılması sürecinde kontrolün verinin öznesinde kalması veri mahremiyeti açısından önemli bir gereksinimdir. Bu süreçte verilerin hangi amaçla işleneceği, kimlerle paylaşılacağı, nerelere aktarılacağı şeffaf ve kontrol edilebilir düzeyde veri öznesi tarafından kontrol edilebilmelidir. Verilerin toplanması ve paylaşılması ile ilgili temsili mahremiyet koruma süreci Şekil 1'de verilmiştir. Verilerin toplanmasından paylaşılmasına kadar geçen süreçte veri mahremiyeti ile ilgili önlemlerin alınabilmesi amacıyla bu süreç takip eden paragraflarda özetlenmiştir [11].

Veri Öznesi, paylaşılan veriler içerisinde korunması gereken verileri yer alan gerçek kişi, kurum ve kuruluşlardır. Şekil 1'de gösterildiği gibi, veri özneleri farklı amaçlar için gerekli olan verilerini işlenmesi amacıyla elektronik uygulamalar aracılığıyla veri sorumlusuna iletir.



Şekil 1. Mahremiyet Korumalı Veri Paylaşımı (Privacy Preserving Data Sharing)

Veri Sorumlusu / Veri İşleyen, veri ihlallerinin oluşmaması amacıyla mahremiyet önlemlerini alması gereken gerçek veya tüzel kişilerdir. Güvenilir olduğu varsayılan kanuni yükümlülükleri bulunan veri sorumlusu veya veri işleyenler elektronik uygulamalar aracılığıyla toplamış oldukları verileri uygun yöntemlerle depolamakta, kullanmakta ve gerektiğinde anonimleştirerek paylaşmaktadır. Veri sorumluları veya veri işleyenler anonimleştirme sonucunda verilerde meydana gelen kayıpları metrikler ile ölçerek veriden sağlanan faydanın hesaplamasını yaparlar.

Toplanan veriler, ID (doğrudan tanımlayıcı), SA (hassas) ve QID (dolaylı tanımlayıcı) olmak üzere 3 grupta sınıflandırılarak T (ID, QID, SA) biçiminde ifade edilir [12]. Anonimleştirme işlemi sonucunda QID nitelikleri eşit kayıtların oluşturduğu eşdeğer grupları içeren anonimleştirilmiş yeni tablo T* ile gösterilmektedir [13]. Mahremiyet korunmasıyla ilgili yapılan çalışmalar takip eden alt bölümlerde özetlenmiştir.

a. Veri Anonimleştirme (Data Anonymization)

Anonimleştirme, verinin tipi ve biçimi korunarak veri faydası açısından kabul edilebilir düzeyde yapılan mahremiyet koruyucu işlemlerdir. Veri anonimleştirme ilk defa 1981 yılında Chaum tarafından önerilmiş ve ilk uygulama Jakobsson tarafından yapılmış ve literatürde yaygın olarak kullanılan anonimleştirme yöntemleri takip eden paragraflarda özetlenmiştir [14-18].

Genelleştirme (Generalization), QID değerlerinin biçimsel ve anlamsal bütünlüğünü koruyarak daha genel olarak ifade edilmesini sağlar. Sayısal değerler aralıklar içerisinde, kategorik değerler ise üst seviye kapsayıcı bir değer ile genelleştirilir. *Baskılama(Supression),* QID niteliklerinin bazı karakterlerinin veya tamamının belirlenen bir karakterle değiştirilerek anonimleştirilmesi sağlanır. *Kovalara ayırma (Bucketization),* QID nitelikleri ile hassas nitelikler arasındaki ilişkinin zayıflatılmasıyla yapılan anonimleştirme yöntemidir. *Hibrit yöntem (Hybrid method),* genelleştirme ve bastırma yöntemlerinin birlikte kullanılmasıyla yapılan anonimleştirme tekniğidir. Bu yöntemde veri faydası esas alınarak tablonun anonimleştirilmesinde genelleştirme ve bastırma teknikleri birlikte kullanılmaktadır. Hibrid anonimleştirme ile ilgili örnek Tablo 1'de verilmiştir.

Tablo 1. Hibrit anonimleştirme (Hybrid Anonymization)

QID			*SA*
Cinsiyet	*Yaş*	*Ülke*	*Hastalık Durumu*
Erkek	>40	*	Enfarktüs
Erkek	>40	*	Enfeksiyon
Kadın	<40	*	Enfeksiyon
Kadın	<40	*	Enfeksiyon

Tablo 1 incelendiğinde araştırmacı için 40 yaşının önemli olduğu bu yaşın altında veya üzerinde bilgisinin kendisine yeterli olmasından dolayı genelleştirme yöntemiyle yas verisinin

anonimleştirildiği görülmektedir. Ülke niteliğiyle ilgili bilgiye ihtiyacı olmadığı için bu nitelik tamamen baskılanmıştır. Veri alıcısının ihtiyacı olan fayda çerçevesinde ilgili veriler hibrit yöntemlerle anonimleştirilerek yayınlanmıştır.

b. Veri Madenciliğinde Mahremiyetin Korunması (Privacy Preserving Data Mining-PPDM)

PPDM, veri sahiplerinin kimliklerinin veya hassas bilgilerinin ifşa edilmesini engelleyen veri kümesi üzerinde birden fazla araştırmacının birlikte çalışmasını mümkün kılan mahremiyet korumalı veri madencilik yöntemidir [19-22]. Lindell, çok taraflı bir ortamda hassas ve gizli bilgiler içeren veritabanlarından elde edilecek veri kümeleri üzerinde her bir tarafın verisinin diğer taraflara ifşa edilmeden veri madenciliği işlemi yapılmasını mahremiyeti koruyan veri madenciliği olarak tanımlamıştır [23]. Veri madenciliğinde mahremiyetin korunmasını sağlamak amacıyla istatiksel ve kriptografik temelli farklı yaklaşımlar geliştirilmiştir. [24]. PPDM ile ilgili yaklaşımlar takip eden paragraflarda özetlenmiştir.

Verilerin Bozulması (Data Perturbation), verilerin istatistiksel bütünlüğü korunarak anlamlı olarak bozulmasıyla mahremiyet saldırılarına karşı dayanıklı verilerin oluşturulması amaçlanmaktadır [25-27]. Veri bozma işleminde orijinal verilere anlamlı gürültülerin eklenmesi yaygın olarak kullanılmaktadır [28-30]. Gürültü ekleyerek veri bozmanın mahremiyeti korumada etkin olamadığı, yeniden dağılımın oluşturulması ile ilgili saldırılara karşı zayıflıkları olduğu gösterilmiştir [31,32]. Gürültü ekleme yönteminin mahremiyet korumadaki zafiyetlerini gidermek için eklemeden farklı olarak çarpma yöntemi ile ilgili çalışmalar yapılmıştır. Ancak çarpma yönteminin de benzer saldırılara karşı korunmasız olduğu görülmüştür [33,34]. Veri bozmada bir diğer yaklaşım ise kümelenme yöntemidir [35-37]. Veri bozma yaklaşımlarının veri faydası üzerinde olumsuz etkisi olduğundan ve saldırılara karşı dirençli olmadığından fayda temelli veri modellerinde çoğunlukla tercih edilmemektedir.

Çok Taraflı Güvenli Hesaplama (Secure Multi-party Computation-SMC), çok taraflı güvenli hesaplama (ÇGH), veri paylaşımında bulunacak tarafların mahrem verilerinin ifşa edilmeden, taraflarca belirlenen ortak bir fonksiyon aracılığıyla polinom zamanda güvenli hesaplama yapılmasını sağlayan kriptografik bir yöntemdir [38,39]. Literatürde ÇGH ile ilgili birçok çalışma yapılmıştır [40-46].

c. Veri Yayınlanmasında Mahremiyetin Korunması (Privacy Preserving Data Publishing-PPDP)

Veri yayınlama yöntemi, veri paylaşımında kolay ve ekonomik bir yöntem olduğu için veri sorumluları veya veri işleyenler tarafından yaygın olarak kullanılır. Literatürde veri yayınlamada yaygın olarak kullanılan mahremiyet modelleri takip eden paragraflarda örnekler verilerek özetlenmiştir [12,48,49].

k-Anonimlik modeli (k-Anonimity model), verilerin anonimleştirilmesinde Samarati ve Sweeney tarafından önerilen k-anonimlik modeli verilerin yayınlanmasında en yaygın kullanılan veri öznesinin kimliğini koruyan mahremiyet modellerinin başında gelmektedir. [1,4].

Tablo 2. 4-Anonim tablo (4-Anonymous table)

	QID			*SA*
	P.Kodu	*Yaş*	*Uyruk*	*Hastalık Durumu*
1	130**	≤30	*	Kalp Hastalığı
4	130**	≤30	*	Kalp Hastalığı
9	130**	≤30	*	Viral Enfeksiyon
10	130**	≤30	*	Viral Enfeksiyon
5	1485*	≥40	*	Kanser
6	1485*	≥40	*	Kalp Hastalığı
7	1485*	≥40	*	Viral Enfeksiyon
8	1485*	≥40	*	Viral Enfeksiyon
2	130**	3*	*	Kanser
3	130**	3*	*	Kanser
11	130**	3*	*	Kanser
12	130**	3*	*	Kanser

ID niteliklerin tablodan çıkarılmasından sonra, yayınlanacak tabloda en az k adet kaydın QID değerinin aynı olduğunu garantilemektedir. Yayınlanan tablodaki her bir kayıt diğer k-1 kayıtla QID nitelikleri aynı olduğundan kimlik ifşasının engellendiği k-anonim bir tablo meydana gelmektedir. İlk bakışta basit bir problem olarak görünmesine karşılık optimum k-anonimliği sağlamanın NP-Hard bir problem olduğu ispatlanmış ve yaklaşık çözümler üretilmeye çalışılmıştır [1,13,50]. QID (P.Kodu, Yaş, Uyruk) ve SA (Hastalık Durumu) olan

tabloda QID niteliklerinden P.Kodu'nun iki hanesi baskılnarak gizlenmiş, yaş genelleştirilmiş ve uyruk tamamen baskılanarak P.Kodu ve Yaş'a göre 4-Anonim çizelgesi Tablo 2'de verilmiştir.

k-Anonimlik modeli kimliklerin ifşa edilmesine yönelik saldırılara karşı koruma sağlarken niteliklerin ifşa edilmesine yönelik saldırılara karşı koruma sağlayamamaktadır. Örneğin, Tablo 2'de verilen anonim tabloda posta kutusu 130 ile başlayan 30'lu yaşlardaki kişi sayısı 4 olmasına rağmen, hastalıkların hepsinin aynı olması bu gruba ait veri sahiplerinin hastalık niteliğinin ifşa edilmesine sebep olmaktadır. Literatürde, verilerin yapısal özelliklerine göre k-Anonimlik yaklaşımından türetilmiş, k-derece anonimlik [51], k-aday anonimlik [52], k-otomorfizm anonimlik [53], k-komşu anonimliği [54,55] ve k, l-gruplama [56] gibi farklı modeller sunulmuştur.

*l-Çeşitlilik modeli (l-Diversity model)***,** Machanavajjhala ve arkadaşları k-Anonimlik modelinin zayıflıklarını göstererek anonimleştirilmiş grupların homojen değerlerden oluşmasını engellyen l-Çeşitlilik modelini önermişlerdir [13]. 3-çeşitli anonim yapı Tablo 3'de verilmiştir.

Tablo 3. 4-Anonim, 3-Çeşitli tablo (4-Anonymous, 3-Diversity table)

	QID			*SA*
	P.Kodu	*Yaş*	*Uyruk*	*Hastalık Durumu*
1	1305*	≤40	*	Kalp Hastalığı
4	1305*	≤40	*	Viral Enfeksiyon
9	1305*	≤40	*	Kanser
10	1305*	≤40	*	Kanser
5	1485*	≥40	*	Kanser
6	1485*	≥40	*	Kalp Hastalığı
7	1485*	≥40	*	Viral Enfeksiyon
8	1485*	≥40	*	Viral Enfeksiyon
2	1306*	≤40	*	Kalp Hastalığı
3	1306*	≤40	*	Viral Enfeksiyon
11	1306*	≤40	*	Kanser
12	1306*	≤40	*	Kanser

l-Çeşitlilik modeli her bir QID grubu içerisindeki SA niteliklerindeki istenen çeşitliliğini garanti etmesine rağmen değerler arasındaki anlamsal yakınlıklara ve dağılım yığılmalarına dayalı olarak yapılan saldırılara karşı koruma sağlayamamaktadır.

t-Yakınlık modeli (t-Closness model), l-Çeşitlilik yaklaşımının yeterli korumayı sağlayamaması üzerine Li ve arkadaşları tarafından t-Yakınlık modeli önerilmiştir [57]. Önerilen bu model her bir QID grubu içerisindeki SA niteliklerin anlamsal yakınlıklarının dengelenebilmesi amacıyla birbirlerine t-Yakınlıkta olmasını garanti etmektedir. Buna göre her bir eşdeğer grup içerisindeki hassas nitelik dağılımının tüm tablo içerisindeki niteliklerin dağılımına olan uzaklığı t ile belirlenen eşik değerini geçmeyecektir. t-Yakınlık yaklaşımı niteliklerin ifşasına karşı koruma sağlarken kimliklerin açığa çıkarılmasına karşı koruma sağlayamamaktadır. Mahremiyetin korunmasında nitelik ve kimlik ifşasına yönelik saldırılara karşı koruma sağlanması için t-yakınlık ve k-anonimlik yöntemleri birlikte kullanılmaktadır [58].

Diferansiyel mahremiyet (Differantial privacy), k-Anonimlik, l-Çeşitlilik ve t-Yakınlık yaklaşımları verilerin tamamının mahremiyetini korumaya çalışan bütünsel yaklaşımlardır. Diferansiyel mahremiyet bu yaklaşımlardan farklı olarak kayıt bazında verilerin mahremiyetini korumaya çalışan kısmi bir yaklaşımdır. Veritabanı sorgu sonuçlarının mahremiyetini korumak için Dwork tarafından önerilmiştir [59,60]. Bu modelde sorguların veritabanına gönderilmesi ile veritabanının sorguya cevap vermesi arasında meydana gelecek saldırılara karşı mahremiyetin korunması hedeflenmektedir. D_1 veritabanının içerdiği tekil bir kaydı D_2 veritabanının içermediği bir durumda, ortalama maaş gibi bir sorgunun iki veri kümesi için de sonuçlarının $\exp(\varepsilon)$ 'den farklı olamayacağını garanti etmektedir. İki sorgunun sonuçları arasındaki farkın ihmal edilebilecek kadar küçük olması, kötü niyetli kişilerin sorgu sonucunun D_1'den mi yoksa D_2'den mi geldiğini tahmin edememesini sağlayacaktır. Böylece iki veri tabanı arasındaki tekil bir kaydın varlığı ortaya çıkmamış olacaktır.

d. Veri Toplanmasında Mahremiyetin Korunması (Privacy Preserving Data Collecting-PPDC)

Üçüncü tarafların doğrudan veri sahiplerinden veri toplama ihtiyacı olduğu durumlarda mahremiyetin korunması daha önemli hale gelmektedir. PPDC ile ilgili literatürde yapılan çoğu çalışmada veriler üzerinde bozma algoritmaları kullanarak elde edilen anonim

verilerin veri sahipleri tarafından doğrudan veri toplayıcıya gönderildiği C2S (client to user) mimarisinin önerildiği görülmüştür [61-63]. Zhang ve arkadaşları bu mimarinin geliştirilmiş hali olan hem veri sahiplerinin hem de veri toplayıcının güvendiği sunucu üzerinde verilerin toplanmasını sağlayan CS2U (Client server to user) modelini önermişlerdir [63]. Bu modelde veriler sunucuda toplanmakta ve genelleştirilerek sıkıştırılmış olarak güvenilir sunucuya gönderilmektedir. Xue ve arkadaşları sunucu istemci mimarisinde çalışan kriptografik yöntemleri kullanan bir PPDC modeli önermişlerdir [64]. Bu modelde veri sahipleri şifreli olarak mesajlarını sunucuya iletmekte sunucuda toplanan veriler üzerinde gerekli işlemler çok taraflı güvenli hesaplama ile güvenilmeyen sunucu üzerinde yapılarak veriler k-anonim olarak veri toplayıcıya sunulmaktadır. Williams ve arkadaşları veri sahipleri tarafından seçilecek mahremiyet düzeyine göre hiyerarşik temelli bir PPDC yaklaşımı önermişlerdir [65].

Bu bölümde özetlenen önlemlerin yeterli olmaması, kullanılan modellerin zafiyetleri veya diğer sebeplere bağlı olarak saldırganlar tarafından mehremiyet saldırıları gerçekleştirilmektedir. Mahremiyet saldırılarının bilinmesi mevcut yöntem ve modellerin zayıflıkları ile aksayan yönlerini ortaya çıkararak yeni yaklaşımların sunulması açısından önem taşımaktadır. Bu amaçla takip eden bölümde mevcut yöntemleri ve modelleri hedef alan mahremiyet saldırıları özetlenerek değerlendirmeler yapılmıştır.

3. MAHREMİYETE KARŞI SALDIRILAR (ATTACKS TO PRIVACY)

Bu bölümde mahremiyet modellerinin mevcut zafiyetlerini ortaya çıkaran ve yeni modellerin gelişmesine katkıda bulunan yaygın saldırı yöntemleri incelenmiştir. Literatürde mahremiyet ihlallerine yol açan saldırılardan önemli olanları takip eden alt başlıklarda özetlenmiştir.

a. Arkaplan Bilgi Saldırıları (Background Knowledge Attacks)

Saldırganlar tarafından farklı yollardan elde edilen arka plan bilgileri mahremiyet saldırılarının ve ihlallerinin yaşanmasında önemli rol oynamaktadır. Arka plan bilgileri farklı kuruluşlar tarafından yayınlanan verilerden, sosyal ağlardan, gazete ve dergilerden, gerçek dünyadaki sosyal ilişkilerden ve benzeri diğer yollardan elde edilebilen hassas olmayan bilgilerdir. Veri bağlama yöntemleriyle arka plan bilgilerinin başka kayıtlarla ilişkilendirilmesi sonucunda mahremiyet ihlalleri yaşanmaktadır [66]. Arka plan bilgilerinin saldırılarda kullanılabilmelerini engellemek zordur. Veri sahiplerinden tanıtım, kampanya, araştırma vb. gibi taleplerle istenen bilgilerin, arka plan bilgileriyle bir araya getirilip yorumlandığında mahremiyet ihlaline yol açacağı her zaman akılda tutulmalıdır.

b. Homojenlik Saldırıları (Homogenity Attacks)

Anonim tablolar içerisinde yer alan gruplar içerisindeki hassas niteliklerin tamamı veya çoğunluğunun benzer olduğu durumlarda veri sahipleri homojenlik saldırılarına karşı korunmasız kalmaktadır. Homojenlik saldırılarının engellenebilmesi amacıyla anonim tablodaki gruplar içerisindeki benzer hassas niteliklerin aynı grup içerisinde yer almasının engellenmesi veya bu çalışma kapsamında önerilen kayıt çoğaltılma yaklaşımının kullanılması gerekmektedir. Kayıt çoğaltma yaklaşımıyla homojen olan nitelikler seyreltilerek heterojen hale getirilecek ve ihlal riski azaltılacaktır.

c. Çarpıklık Saldırıları (Skewness Attacks)

Yayınlanan veya paylaşılan anonim veri kümeleri içerisinde SA değerlerinin istatistiki dağılımları mahremiyetin sağlanması açısından önemlidir. Hassas niteliklere ait değerlerin bir kısmının çok baskın olması dağılımları olumsuz yönde etkilemekte ve hassas niteliklerin genel dağılımında çarpıklık meydana getirmektedir. SA nitelik değerlerinin genel dağılımları çarpık olduğunda yayınlanan veya paylaşılan veri kümeleri çarpıklık saldırılarına karşı korunmasız olacaktır [67].

d. Anlamsal Benzerlik Saldırıları (Similarity Attacks)

Anonim gruplar içerisindeki SA nitelik değerlerinin sadece birbirinden farklı olması mahremiyetin korunması açısından yeterli olmamaktadır. SA nitelik değerlerinin sezgisel benzerliklerinden faydalanılarak yapılan anlamsal benzerlik saldırıları ve homojenlik

saldırıları birbirine benzemektedir [68]. Anlamsal benzerlik saldırılarının engellenebilmesi amacıyla aynı anonim grupta yer alan hassas niteliklerin benzerliklerinin hesaplanarak benzer SA nitelik değerlerinin farklı gruplarda yer almasını sağlayacak t-yakınlık ve benzeri çözümlere ihtiyaç vardır.

e. Minimalite Saldırıları (Minimality Attacks)

Wong ve arkadaşları saldırganların anonimleştirme algoritmaları veya mekanizmaları hakkında bilgi sahibi olduğu durumlarda minimalite saldırılarını yapabileceklerini göstermişlerdir [69]. Minimalite saldırıları minimizasyon prensibi üzerine kurulmuştur. Bu prensibe göre anonimleştirme süreçlerinde veri üzerindeki değişiklikler minumum seviyede kalmalı ve gereğinden fazla anonimleştirme yapılmamalıdır [69].

f. De-Finetti Saldırıları (De-Finetti Attacks)

Kifer, Definetti teoremi ve değiştirilebilirlik kavramları birlikte kullanılarak mahremiyet hakkında çıkarım yapılabileceğini teorik ve deneysel yöntemlerle göstermiştir [70,71]. Diğer saldırılarla deFinetti saldırısı arasındaki en önemli fark saldırganın kapsamlı bir arka plan bilgisine ihtiyaç duymamasıdır. Veri kümesi içerisindeki ilgili kaydın hassas olmayan nitelikleri üzerinde makine öğrenme modeli kurularak deFinetti saldırısı yapılabilmektedir.

4. VERİ FAYDASININ ÖLÇÜMÜ (MEASURING UTILITY OF DATA)

Veri kalitesinin ve faydasının ölçülmesi mahremiyetin korunmasında veri alıcıları açısından çok önemli bir konudur. Mahremiyetin korunması için verilerin anonimleştirilmesinin yanında anonimleştirilmiş verilerin veri alıcılara sağlayacağı faydanın ölçülmesi gerekmektedir.

Veri faydasını ölçmek için anonimleştirilmiş tablo ile orijinal tablo arasındaki veri kaybının ölçülmesini sağlayan metriklere ihtiyaç duyulmaktadır [1,4]. Bu tür senaryolarda en makul ölçüm anonimleştirilmiş veri ile orijinal veri arasında benzerliğin ölçülmesidir [72]. Anonimleştirilmiş verinin faydası alıcıların veri ihtiyaçlarına göre değişmektedir. Web üzerinden yayınlanan bir veriye farklı alıcılar farklı amaçlar için eriştiklerinden farklı veri faydası elde edeceklerdir. Bundan dolayı belirlenen bir ilgi metriği bir veri alıcısı için iyi sonuçlar verirken diğer veri alıcısı için aynı sonuçları vermeyebilir.

a. Minimal Bozulma (Minimal Distortion)

Minimal bozulma yöntemine göre veri faydasının ölçümünde her bir niteliğe ait genelleştirilen örnek sayısı kadar ceza puanı uygulanmasıdır. Uygulanan ceza puanlarının toplamı ilgili nitelik için anonimleştirme sırasında meydana gelen bilgi kaybının ölçülmesini sağlar. Veri fayda metriği olarak literatürde birçok çalışmada kullanılmıştır [1,4,73,74].

b. I-Loss

I-Loss kategorik nitelikler için spesifik bir değerin daha genel değere dönüştürülerek anonimleştirilmesiyle meydana gelen veri kaybının ölçülmesi amacıyla Iyengar tarafından önerilmiştir [75]. Bu metriğe göre veri faydası $|V_g|$, ilgili düğüme ait çocukların sayısı, $|DA|$ ise V_g'nin A niteliğindeki değerlerinin sayısı olmak üzere $ILoss(v_g) = |V_g|-1/|DA|$ formülüyle hesaplanır

c. Ayırt Edilebilirlik Yöntemi (Discernibility Method)

Ayırt edilebilirlik metriği (DM), QID dikkate alınarak her bir kaydın diğer kayıtlardan farklılığına göre ceza puanı verilmesi esasına göre veri faydasını ölçmektedir [76]. DM yönteminde, eğer bir kayıt s büyüklüğünde bir gruba ait ise bu kayıt için ceza puanı s, ilgili grup için s^2 ve genelleştirilmiş tablo için $DM(T)=\Sigma(S_i)^2$ olarak hesaplanacaktır.

d. Sınıflandırma Yöntemi (Classification Methods)

Sınıflandırıcıların eğitilmesi amacıyla yayınlanan anonim verilerin anonimleştirme kayıplarının ölçülmesini sağlayan sınıflandırma yöntemi Iyengar tarafından önerilmiştir [75]. Yayınlanan veriler hem sınıflandırma modelini geliştirebilecek yararlı bilgileri hem de sınıflandırma modelini bozabilecek faydasız gürültüleri içermektedir. Yararlı sınıflandırma bilgisi sınıflandırmanın elde edilmesine yardımcı

olduğundan veri yayınlanmadan önce yararsız gürültülerin anonimleştirme yoluyla giderilmesi gerekmektedir.

e. Trade-Off Yöntemi (Trade-Off Methods)

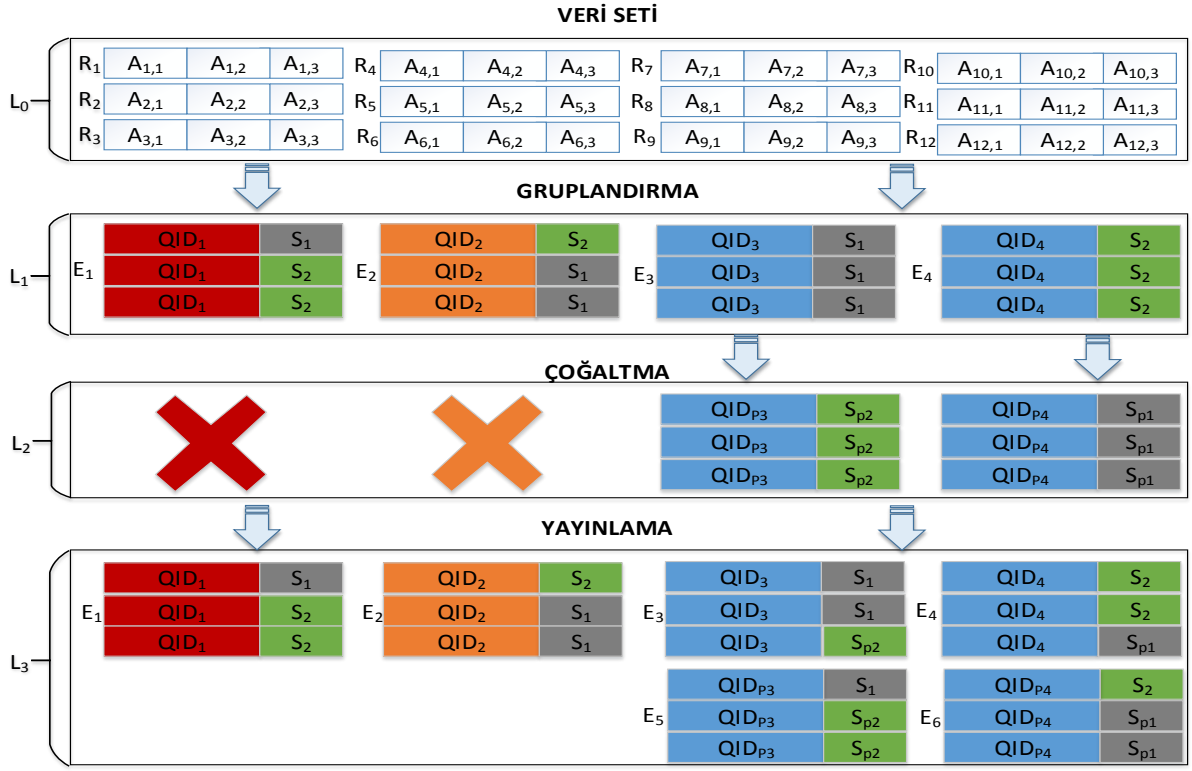
Fung ve arkadaşları bilgi kazancı/mahremiyet kaybı arasındaki dengeyi gözeten veri faydası metriğini önermişlerdir [77,78]. Trade-Off yöntemi anonimleştirme sürecinde mahremiyet ve veri faydasını dikkate alarak iki gereksinim arasındaki dengenin kurulmasına odaklanmaktadır.

5. YENİ YAKLAŞIM: ÇOĞALTMA TEKNİĞİ İLE VERİ MAHREMİYETİNİN SAĞLANMASI (NEW APPROACH: PRIVACY PRESERVING WITH DATA POPULATING METHOD)

Kimlik ve nitelik ifşasına yönelik mahremiyet saldırıları ve bu saldırıları engellemeye yönelik mahremiyet koruyucu yöntemler önceki bölümlerde özetlenmiştir. Mahremiyet koruyucu yöntemlerden k-Anonimlik modeli kimlik ifşasına, l-Çeşitlilik ve t-Yakınlık yöntemleri ise nitelik ifşasına karşı koruma sağladığı önceki bölümlerde örneklerle gösterilmiştir. Bu çalışma kapsamında hem nitelik hem kimlik ifşasına karşı çoğaltma yöntemiyle koruma sağlayan Şekil 2'de temsili olarak gösterilen yeni bir yaklaşım önerilmiştir.

Şekil 2'de görüldüğü üzere 4 katmanlı (L_0-L_4) yeni bir yaklaşım önerilmiş ve bu yaklaşım maddeler halinde aşağıda açıklanmıştır.

1) L_0 katmanında orijinal veriler yer almakta ve aşağıdaki işlemler yapılmaktadır.
   a) Niteliklerin sınıflandırılması
   b) ID niteliklerin veri setinden çıkarılması.
   c) QID niteliklerinin belirlenmesi
2) L_1 Gruplandırma katmanında eşdeğer gruplar oluşturularak aşağıdaki işlemler yapılmaktadır
   a) Eşdeğer gruplar içerisinde yer alacak kayıt sayısının seçilmesi
   b) QID niteliklerin anonimleştirilmesi
   c) Eşdeğer grupların oluşturulması
   d) Eşdeğer gruplara mahremiyet taraması yapılması
   e) Zafiyet ve zayıflık tespit edilen grupların çoğaltma katmanına gönderilmesi



Şekil 2. Veri çoğaltma yaklaşımı (Data populating approach)

3) L_2 Çoğaltma katmanında veri faydasını gözeterek mahremiyeti koruyan önlemler alınmakta ve aşağıdaki işlemler yapılmaktadır.
   a) Güncel mahremiyet saldırıları (homojenlik, benzerlik, dağılım yığılmaları vb.) dikkate alınarak çoğaltma yapılacak veriler oluşturulur.
   b) Çoğaltılan mahremiyet koruyucu gerçek olmayan veriler zafiyet ve zayıflık içeren gruplarla ilişkilendirilir.
4) L_3 Yayınlama katmanında aşağıdaki işlemler yapılarak veriler yayınlanmaktadır.
   a) Çoğaltma katmanında oluşturulan gerçek olmayan veriler ilişkili gruplara yerleştirilir.
   b) Çoğaltma kayıtlarıyla oluşturulan gruplar ile diğer gruplar birleştirilerek veriler yayınlanır.

Önerilen yaklaşımda kullanılan nostasyonlar aşağıdaki Tablo 4'de verilmiştir.

Tablo 4. Notasyonlar (Notations)

Notasyon	Açıklama
R_i	i. kayıt
$A_{i,j}$	i. kaydın j. Niteliği
L_k	k. katman
E_m	m. eşdeğer grup
QID_m	m. QID niteliği
S_n	n. SA niteliği
QID_{pm}	m. çoğaltılan QID niteliği
S_{pn}	n. çoğaltılan SA niteliği
K_s	Kayıt sayısı

Şekil 2'de verilen temsili gösterimde 12 kayıttan (R_1-R_{12}) oluşan 3 niteliğe sahip ($A_{i,j}$) veri seti örnek olarak alınmıştır. Gruplar içinde yer alan kayıt sayısının K_s=3 olarak seçilmesiyle kimlik ifşa riski %33 olarak hesaplanmakta ve bu oranın kimlik ifşa koruması için yeterli olduğu varsayılmaktadır. SA niteliğinin iki değerden birisini alabilme ihtimalinden dolayı nitelik ifşa riskinin %50 olması gerekmektedir. Aksi durumda aynı grup içerisinde SA değerleri eşit olan homojen kayıtlar oluşacağından nitelik ifşası olacaktır. Hassas değerlerin homojen olduğu gruplarda mahremiyet ihlallerini engellemek için çoğaltma işlemi yapılması gerekmektedir.

Örnekte, QID üzerinde anonimleştirme işlemleri yapılarak 4 tane eşedeğer grup (E_1-E_4) oluşturulmuştur. Oluşan gruplar incelendiğinde E_3-E_4 gruplarının hassas niteliklerinin homojen durumda olduğu tespit edilir. Nitelik ifşasının engellenebilmesi amacıyla bu gruplar için çoğaltma katmanında işlem yapılması gerekir. Çoğaltma katmanında E_3-E_4 gruplarının zafiyetini gidermek için E_{p5}-E_{p6} grupları oluşturularak ilgili gruplarla ilişkilendirilir. Gerekli çoğaltmalar sonucunda, tüm gruplarda kimlik ifşa oranının %33 hassas nitelik ifşa oranının %50 olduğu görülmekte ve bu durum mahremiyet koruması açısından kabul edilebilir bir durumdur.

Hassas nitelikler açısından veri faydası incelendiğinde, orijinal veri setinde S_1 ve S_2 değerlerinin genele oranının 6/12 olduğu görülür. Çoğaltma işlemi sonrasında ise S_1 ve S_2 değerlerinin oranının 9/18 olduğu ve oranın matematiksel açıdan değişmediği gözlemlenmiştir.

Önerilen yaklaşımının temelinde veri faydasını gözeterek mahremiyet koruyucu gerçek olmayan verilerin orijinal verilere eklenmesiyle ihlal riskinin azaltılması hedeflenmektedir. Veri çoğaltmanın mahremiyet korumasına sağladığı faydanın matematiksel ifadesi aşağıda verilmiştir.

Orijinal tablo T, yayınlanacak tablo T^*, T_0 ise $|T^*|=|T_0|$ olmak üzere rasgele seçilmiş bir tablo olduğunda, T^* tablosunda s_i niteliğinin yer aldığı R kaydının T_0 tablosunda yer alma olasılığı, $P(S(R)=s_i \backslash T_0)$ olarak gösterilir. Bu olasılığı, P(R)'nin başarı olasılığı ve $n=|T_0|$ olacak şekilde binom dağılımıyla aşağıdaki şekilde ifade etmek mümkün olup

$$P(S(R) = s_i \backslash T_0) = 1\text{-}f(0;n,P(R) = 1\text{-}(1\text{-}P(R))^n$$

denklemiyle ifade edilir.

Örneğin, R(20-30, erkek, prostat kanseri) kaydı için P(R)=0,00175 ve n=100 değerleri verilmektedir. Yukarıda verilen denkleme göre T_0'ın olasılığı 0,16 olacağından her 100 kayıttan 16 tanesinin R kaydıyla aynı olması beklenecektir. Anlamlı kayıt çoğaltma yöntemiyle veri setini genişleterek n=1000 yaptığımızda yine aynı denkleme göre olasılık 0,82 olarak hesaplanacak ve her 100 kayıttan 82 tanesinin R kaydıyla aynı olması beklenecektir. Saldırganın 16 kayıt üzerinden kurbanın kaydına

ulaşabilme ihtimali 1/16=0,06 olurken 82 kayıt içerisinden kurbanın kaydına ulaşma ihtimali 1/82=0,012 olarak bulunacaktır.

Örneklerden elde edilen sonuçlar yorumlandığında veri çoğaltma yaklaşımının veri faydasını gözeterek mahremiyet saldırılarına karşı verileri daha dirençli hale getirdiği gözlemlenmiştir. Önümüzdeki çalışmalarda yüksek sayıda kayıtlar içeren veri kümeleri üzerinde deneyler yapılarak önerilen yöntemin performansının boyut, yük, işlem gücü gibi parametreler üzerinden tartışılarak diğer yöntemlerle karşılaştırılması gerekmektedir. Ayrıca ilişkisel veriler dışında kalan diğer veri tipleri üzerinde veri çoğaltma yaklaşımının uygulanabilmesi amacıyla çalışmalar yapılması gerekmektedir.

6. SONUÇLAR ve DEĞERLENDİRMELER (RESULTS and CONCLUSIONS)

Veri mahremiyetinin korunmasıyla ilgili gerekli teknik ve idari tedbirlerin alınmaması veya yetersiz olmasına bağlı olarak hassas ve kişisel verilerin ihlal edilmesiyle ortaya çıkan mahremiyet problemleri her geçen gün çeşitlenerek artmakta ve maddi ve manevi maliyetlere yol açmaktadır. Mahremiyet ihlallerinin oluşmaması için veri yaşam döngüsünde mahremiyetin korunması amacıyla etkili ve yeterli önlemlerin alınması gerekmektedir. Bu çalışmada veri mahremiyetinin korunmasına yönelik kapsamlı bir çalışma yapılmış, yapılan çalışma, tespitler, değerlendirmeler ve öneriler aşağıda maddeler halinde verilmiştir.

- Çalışma kapsamında yapılan araştırmalar, yaşanan mahremiyet ihlalleri, uzman personel eksikliği, yanlış yapılan yatırımlar, yaşanan veri ihlalleri dikkate alındığında ülkemizde kişisel verilerin korunması konusunun hassasiyeti ve öneminin arttırılması amacıyla farkındalık çalışmalarının yapılması gerekliliği tespit edilmiştir.
- Veri mahremiyeti konusunda yeterli birikim ve tecrübelerin oluşmadığı ülkemizde veri mahremiyetine dayalı yerli çözümlerin ve bu alanda çalışan firmaların az olmasından anlaşılmaktadır.
- Ülkemizde veri mahremiyeti konusunda yapılan araştırmaların, ar-ge projelerin ve yayınların yetersiz olduğu bu alanda çalışan akademisyen sayısı ve akademik çalışmaların az olmasından anlaşılmaktadır.
- Veri mahremiyet ihlallerinin dünyada ve ülkemizde her geçen gün arttığı yaşanan ihlallerinin ülkelerin güvenliğini ve özellikle mahremiyetini tehdit eden boyutlara ulaştığı gözlemlenmiştir.
- Mahremiyet tehditlerinin ülkeden ülkeye, toplumdan topluma, kişiden kişiye değişiklik göstermesi dikkate alındığında ülkemize yönelik mahremiyet ihlalleri ve tehditleri konusunda üniversitelerin yeni çözümler geliştirmesi, sektörün de ülkemize özel çözümler üretmesi gerekmektedir.
- Bu çalışma kapsamında mahremiyetin korunmasında veri faydasını gözeten yeni bir veri çoğaltma yaklaşımı sunulmuş ve Bölüm 5 de detaylandırılmıştır. Yeni yaklaşım, yayınlanacak veri kümesini eşdeğer gruplara ayırarak zafiyet taraması yapmaktadır. Veri çoğaltma yöntemiyle zafiyet içeren gruplara yapılan eklemeler sayesinde ifşa riskleri kabul edilebilir sınırlara çekilerek verilerin mahremiyet saldırılarına karşı dirençli hale getirilmesi düşünülmüştür. Önerilen yeni yaklaşımda;
  - sadece zayıflık ve zafiyet içeren grupların ele alınmasıyla, üzerinde işlem yapılacak veri kümesinin minimize edilmesi performans ve veri faydasını olumlu yönde etkilemektedir.
  - bahsedilen yöntemler kullanılarak mahremiyet zafiyetlerinin taranmasını sağlayan, zayıflıkların tespit edilerek etkisine göre sınıflandırılmasını sağlayan mahremiyet zafiyet tarayıcı yazılımlara ihtiyaç duyulduğu görülmüştür.
  - mahremiyet zafiyetlerinin ve zayıflıklarının taranarak veri ihlalleri oluşmadan önce mahremiyet saldırılarına hazırlıklı olunması diğer yöntemlere göre mahremiyetin korunmasında etkin bir sonuç sağlamaktadır.
- Mahremiyet korunmasına etki eden unsurlar içerisinde en zayıf halka olarak adlandırılan insan faktörünün olumsuzluğunu azaltmak amacıyla, özelikle veriler üzerinde işlem yapan kamu ve özel sektör çalışanlarına mahremiyet farkındalığını arttırıcı eğitimler verilmelidir.
- Kişisel verilerin korunması konusunda yeterli farkındalığı olmayan kurum/kuruluş/şirket personellerine eğitim

verilmesi kişisel verilerin korunması açısından önemlidir.

- İlgili kurumlar ve üniversitelerin yapacakları ortak çalışmalarda, kişisel verilerin korunması konusu teknik, idari ve kültür yönleriyle ele alınarak mevcut çözümlerin yileştirilmesi veya yeni çözümlerin üretilmesi gerekmektedir.
- Veri bağlantı teknikleriyle yapılan saldırılar özellikle son dönemde yaşanan veri ihlallerinden sonra çok ciddi bir boyut kazanmıştır. Veri ihlalleri sonucunda kişileri doğrudan veya dolaylı olarak tanımlayan kişisel verilerin olması, başta verisi işlenen gereçek kişileri sonrasında ise veri sorumlularını tehdit etmektedir.
- Veri sorumluları tarafından kişisel veriler 6698 sayılı Kişisel Verilerin Korunması kanununa uygun olarak işlenmeli ve bu kanun kapsamındaki yükümlülüklerini yerine getirmelidir.
- Mahremiyet saldırılarına karşı daha dayanıklı modellerin oluşturulabilmesi için mahremiyet koruma önlemlerinin hem yazılım geliştirme hem de iş süreçlerinin bir parçası olacak şekilde kurumsal bir kültür olarak ele alınması gerekmektedir.
- Veri mahremiyetinin hangi düzeyde korunduğunu gösteren sertifikasyonların oluşturularak veri sahiplerine verilerinin korunduğuna dair tarafsız kurumlarca güvence verilmesi gerekmektedir.
- Veri mahremiyetinin korunmasına yönelik verinin toplanmasından paylaşılmasına kadar geçen tüm süreçlerde uçtan uca veri ekonomisini de dikkate alan yeni çatı çözümlere ihtiyaç vardır.
- Kişisel verilerin yoğun olarak işlendiği mobil cihazlar üzerinde yaygın şekilde kullanılan mobil uygulamalara yönelik mahremiyet önlemlerinin alınmasını sağlayacak yeni yaklaşımlar geliştirilmelidir.
- Veri mahremiyeti ihlallerinin tespit edilmesi, zafiyetlerin incelenmesi, saldırıların analiz edilmesi amacıyla mahremiyet laboratuvarlarının kurularak, mahremiyet araştırmalarının önünün açılması ülkemizde yapılacak yeni araştırmalara ve çözümlere önemli derecede katkı sağlayacaktır.

Sonuç olarak, yukarıda yapılan tespitler, değerlendirmeler ve önerilere ek olarak bu çalışma kapsamında önerilen yaklaşımın mevcut çözümlerle entegre edilerek özellikle diğer yöntemlerle paylaşılan verilerin zafiyet ve zayıflıklarının önceden tespit edilmesiyle veri mahremiyetinin korunmasına önemli katkılar sağlayacağı değerlendirilmektedir.

Gelecek çalışmalarda bu çalışma kapsamında önerilen yöntemin büyük veriler üzerinde yapılacak çalışmalarla desteklenerek mahremiyet riskleri ve veri faydası açısından kıyaslanarak etkinliği ve güvenilirliğini ortaya koyacak yeni çalışmalara ihtiyaç vardır.

KAYNAKLAR (REFERENCES)

[1]. Samarati, P., "Protecting respondent's privacy in micro data release", **IEEE Transaction on Knowledge and Data Engineering**, Cilt 13, No 6, 1010-1027, 2001.

[2]. Korolova, A., Protecting privacy while mining and sharing user data, Doktora Tezi, Stanford Üniversitesi, Bilgisayar Mühendisliği Bölümü, 2012.

[3]. Verykios, S.V., Bertino, E., Fovino, N.I., Provenza, P.L., Saygin, Y., Theodoridis, Y., "State-of-the-art in Privacy Preserving Data Mining", **ACM SIGMOD Record**, Cilt 33, Sayı 1, 50-57, 2004.

[4]. Sweeney, L., "k-Anonymity: A model for protecting privacy," **International Journal of Uncertainty Fuzziness and Knowledge-Based Systems**, Cilt 10, Sayı 5, 557–570, 2002.

[5]. Sweeney, L., "Computational Disclosure Control for Medical Microdata: The Datafly System", **Proceedings of an International Workshop and Exposition**, Washington DC, ABD, 442-453, 1997.

[6]. İnternet: President's Information Technology Advisory Committee, "Revolutionizing health care through information technology", http://www. itrd.gov/pitac /meetings/2004/20040617/20040615_hit.pdf, 2012.

[7]. İnternet: Barbaro, M., Zeller, T., "A Face Is Exposed for AOL Searcher No. 4417749", http://www.nytimes.com/2006/08/09/technology/ 09aol.html? pagewanted=all&_r=0, 2006.

[8]. İnternet: Üstün, G., "e-devlet Skandalı", http://www.milliyet.com.tr/Ekonomi/ HaberDetay.aspx?aType=HaberDetay&Kategori =ekonomi&ArticleID=972537&Date=30.07.2008 &b,200.

[9]. İnternet: "Someone Hacked and Leaked Entire Turkish Citizenship Database Online",

https://www.hackread.com/turkish-citizenship-database-hacked-leaked

[10]. Gökçe, H., Abul, O., "Sensitive knowledge hiding application", E**lectrical, Electronics and Computer Engineering (ELECO),** Bursa,Türkiye, 558-562, 2010.

[11]. Gehrke, J., "Models and Methods for Privacy-Preserving Data Analysis and Publishing ", **The 22nd International Conference on Data Engineering,** Atlanta, ABD, 105-106, 2006.

[12]. Fung, B. C. M., Wang, K., Chen, R., Yu, P. S., "Privacy-preserving data publishing: A survey of recent developments", **ACM Computing Surveys (CSUR),** Cilt 42, Sayı 4, 523-553, 2010.

[13]. Machanavajjhala, A., Gehrke, J., Kifer, D.,Venkitasubramaniam, M., "ℓ-Diversity: Privacy beyond k-anonymity," **The 22nd International Conference on Data Engineering,** Atlanta, USA, 2006.

[14]. Chaum, D.L., "Untraceable electronic mail, return addresses, and digital pseudonyms,", **Communications of the ACM**, Cilt 24, Sayı 2, 84-90, 1981.

[15]. Jakobsson, M., Juels, A., Rivest, R.L., "Making mix nets robust for electronic voting by randomized partial checking, ", **In Proceedings of the 11th USENIX Security Symposium**, San Franscisco, 339-353, 5-9 Agustos 2002.

[16]. Martin, D. J., Kifer, D., Machanavajjhala, A., Gehrke, J., Halpern, J.Y., "Worst-case background knowledge for privacy-preserving data publishing", **ICDE 2007 IEEE 23rd International Conference,** İstanbul, Türkiye, 126-135, 2007.

[17]. Xiao X., Tao Y., "Anatomy: Simple and Effective Privacy Preservation", **Proc. of the 32nd International Conference on Very Large Data Bases,** Seoul, Kore,139−150, 2006.

[18]. Koudas, N., Srivastava, D., Yu, T., Zhang, Q., "Aggregate Query Answering on Anonymized Tables", **ICDE 2007 IEEE 23rd International Conference,** İstanbul, Türkiye, 116-125, 2007.

[19]. Gayatri Nayak and Swagatika Devi (2011), "A Survey On Privacy Preserving Data Mining: Approaches and Techniques", **International Journal of Engineering Science and Technology**, Cilt 3, Sayı 3, 2127-2133, 2011.

[20]. Lindell, Y., Pinkas, B., "Privacy Preserving Data Mining", **20th Annual International Cryptology Conference**, California, USA, 36-53, 2000.

[21]. Hand, D., Mannila, H., Smyth, P., "Principles of DataMining", MIT Press, 2001.

[22]. Vaidya, J., Clifton, C., "Privacy-Preserving Data Mining: Why, How, and When" **IEEE Security & Privacy,** Cilt 2, Sayı 6, 19-27, 2004.

[23]. Du, W., Atallah, M. J., "Secure Multi-Party Computation Problems and Their Applications: A Review and Open Problems", **In Proceedings of New Security Paradigms Workshop**, New Mexico, ABD, 11-20, 2001.

[24]. Adam, N. R., Worthmann, J. C., "Security-control methods for statistical databases: a comparative study" **ACM Computing Surveys (CSUR)**, Cilt 21, Sayı 4, 515–556, 1989.

[25]. T. P. Hong, K. T. Yang, C. W. Lin and S. L. Wang, "Evolutionary privacy preserving data mining", **World Automation Congress (WAC)**, Kobe, Japonya 1-7, 2010.

[26]. Evfimievski, A., Srikant, R., Agrawal, R., Gehrke, J., "Privacy preserving mining of association rules", **Information Systems**, Cilt 29, Sayı 2004, 343-364, 2003.

[27]. Qi, X., Zong, M., "An Overview of Privacy Preserving Data Mining", **Procedia Environmental Sciences**, Cilt 12, Sayı B-2012, 1341-1347, 2011.

[28]. Agrawal, R., Srikant, R., "Privacy Preserving Data Mining", **ACM SIGMOD Record**, Cilt 29, Sayı 2, 439-450, 2000.

[29]. Muralidhar, K., Sarathy, R., "A Theoretical Basis for Perturbation Methods" **Statistics and Computing**, Cilt 13, Sayı 4, 329-335, 2003.

[30]. Evfimievski, A., "Randomization in Privacy Preserving Data Mining" **ACM SIGKDD Explorations Newsletter**, Cilt 4, Sayı 2, 43-48, 2002.

[31]. Kargupta, H., Datta, S., Wang, Q., Sivakumar, K., "On the Privacy Preserving Properties of Random Data Perturbation Techniques", **The Third IEEE International Conference on Data Mining**, Florida, ABD, 99-106, 2003.

[32]. Huang, Z., Du, W., Chen, B., "Deriving Private Information from Randomized Data" **In Proceedings of the 2005 ACM SIGMOD Conference**, Baltimore, ABD, 37-48 2005.

[33]. Kim, J. J., Winkler, W. E., "Multiplicative noise for masking continuous data" **Statistical Research Division U.S. Bureau of the Census**, Washington D.C., ABD, 2003.

[34]. Liu, K., Kargupta, H., Ryan, J., "Random projection-based multiplicative data perturbation for privacy preserving distributed data mining", **IEEE Transactions on Knowledge and Data Engineering (TKDE)**, Cilt 18, Sayı 1, 92–106, 2006.

[35]. Ferrer-Domingo, J., Mateo-Sanz, J.M., "A Comparative Study Of Microaggregation Methods", **Qüestiió Journal**, Cilt. 22, Sayı. 3, 511–526, 1998.

[36]. Ferrer-Domingo, J., Mateo-Sanz, J.M., "Practical data-oriented microaggregation for statistical disclosure control," **IEEE Transactions on Knowledge and Data Engineering (TKDE)**, Cilt 14, Sayı. 1, 189–201, 2002.

[37]. Ferrer-Domingo, J., Torra, V., "Ordinal, continuous and heterogeneous k anonymity through microaggregation," **Data Mining and**

Knowledge Discovery, Cilt. 11, Sayı. 2, 195–212, 2005.

[38]. Yao, A. C., "How to generate and exchange secrets", **In Proceedings 27th IEEE Symposium on Foundations of Computer Science**, Toronto, Kanada, 162–167, 1986.

[39]. Lindell, Y., Pinkas, B., "Secure Multiparty Computation for Privacy-Preserving Data Mining" **The Journal of Privacy and Confidentiality**, Cilt 1, Sayı 1, 59-98, 2009.

[40]. Sheikh, R., Mishra, D.K., Kumar, B., "Secure Multiparty Computation:From Millionaires Problem to Anonymizer" **Information Security Journal: A Global Perspective**, Cilt 20, Sayı 1, 25-33, 2011.

[41]. Yao, A.C., "Protocols for secure computations ", **Proceedings of the 23rd Annual Symposium on Foundations of Computer Science**, Washington, ABD, 160-164, 1982.

[42]. Goldreich, O., Micali, S., Wigderson, A., "How to Play any Mental Game: A Completeness Theorem for Protocols with Honest Majority", **Proc. 19th ACM Symp. Theory of Computing**, New York, ABD, 218–229, 1987.

[43]. Sheikh, R., Kumar, B., Mishra, D. K., "Privacy-Preserving k-Secure Sum Protocol", **In the International Journal of Computer Science and Information Security**, Cilt 6, Sayı 2 , 184-188, 2009.

[44]. Atallah M.J., Kerschbaum, F., Du, W., "Secure and Private Sequence Comparisons," **Proceedings of the 2003 ACM workshop on Privacy in the electronic society**, New York, ABD, 39-44, 2003.

[45]. Goldwasser, S., "Multi party computations: past and present", **Proceedings of the sixteenth annual ACM symposium on Principles of distributed computing**, New York, ABD, 1-6, 1997.

[46]. Maurer, U., "Secure Multi-Party Computation made Simple", **Discrete Applied Mathematics,** Cilt. 154, Sayı 2, 370-381, 2006.

[47]. Yongcheng, L., Jiajin, L., Jian, W., "Survey of Anonymity Techniques for Privacy Preserving", **2009 International Symposium on Computing, Communication and Control (ISCCC 2009)**, Singapur, 248-252, 2011.

[48]. Xiao, X., Tao, Y., "Personalized privacy preservation", **Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data**, New York, ABD, 229–240, 2006.

[49]. Li, H., Ma, J., Fu, S., "Analyzing mechanism-based attacks in privacy-preserving data publishing", **International Journal for Light and Electron Optics**, Cilt 124, Sayı 24, 6939-6945, 2013.

[50]. Samarati, P., Sweeney, L., "Protecting privacy when disclosing information:k-anonymity and its enforcement through generalization and suppression", **SRI International**, Technical Report, SRI-CSL-98-04, 1998.

[51]. Liu, K., Terzi, E., "Towards identity anonymization on graphs", **ACM SIGMOD International Conference on Management of Data (SIGMOD)**, New York, ABD, 93-106, 2008.

[52]. Hay, M., Miklau, G., Jensen, D., Weis, P., Srivastava, S., "Anonymizing social networks", Technical report, University of Massachusetts, 2007.

[53]. Zou, L., Chen, L., Özsu, M.T., "K-automorphism: A general framework for privacy preserving network publication", **Very Large Data Base Endowment**, Cilt 2, Sayı 1, 946-957, 2009.

[54]. Zhou, B., Pei, J., "Preserving privacy in social networks against neighborhood attacks", **IEEE International Conference on Data Engineering (ICDE)**, Cancun, Meksika, 506-515, 2008.

[55]. Wu, X., Ying, X., Liu, K., Chen, L., "A Survey Of Algorithms For Privacy-Preservation Of Graphs And Social Networks", **Managing and Mining Graph Data**, Cilt 40, Editör: Aggarwal, C., Wang, H., Kluwer Academic Publishers, 421-453, 2010.

[56]. Cormode, G., Srivastava, D., Yu, T., Zhang, Q., "Anonymizing bipartite graph data using safe groupings", **The VLDB Journal**, Cilt 19, Sayı 1, 115-139, 2008.

[57]. Li, N., Li, T., Venkatasubramanian, S.,"t-Closeness: Privacy beyond k-anonymity and ℓ-diversity", **In Proceedings of the International Conference on Data Engineering (ICDE)**, İstanbul, Türkiye, 106–115, 2007.

[58]. Rubner, Y., Tomasi, C., Guibas, L. J., "The earth mover's distance as a metric for image retrieval", **International Journal of Computer Vision**, Cilt 40, Sayı 2, 99–121, 2000.

[59]. Dwork, C., "Differential privacy", **Theory and Applications of Models of Computation**, Cilt 4978, Editörler: Agrawal, M., Du, D. Z., Duan, Z., Li, A., Springer Berlin Heidelberg, 1–19, 2008.

[60]. Dwork, C., "Differential Privacy: A Survey of Results", **Theory and Applications of Models of Computation**, Cilt 4052, Editörler: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I., Springer Berlin Heidelberg, 1–12, 2006.

[61]. Du, W., Zhan, Z. "Using randomized response techniques for privacy-preserving data mining", **International Conference on Knowledge Discovery and Data Mining**, San Francisco,, ABD, 505-510, 2003.

[62]. Zhang, N., Wang, S., Zhao, W. "A new scheme on privacy-preserving data classification", **International Conference on Knowledge Discovery and Data Mining**, Chicago, ABD, 374–382, 2005.

[63]. Zhang, L., Zhang, W., "Generalization-based privacy-preserving data collection **International Conference on Data Warehousing and Knowledge Discovery**, Las Vegas, ABD,115–124, 2008.

[64]. Xue, M., Papadimitriou, P., Raissi, C., Kalnis, P., Pung, H.K., "Distributed Privacy Preserving Data Collection using Cryptographic Techniques", Stanford University, Technical Report, 2009.

[65]. Williams, A., Barker, K., "Controlling inference: avoiding p-level reduction during analysis", **Proceedings of the Fifth Australian Symposium on ACSW Frontiers**, Ballarat, Avustralya ,193-200, 2007.

[66]. Chen B., LeFevre, K., Ramakrishnan R., "Privacy skyline: Privacy withmultidimensional adversarial knowledge", **In Proceedings of the International Conference on Very Large Data Bases (VLDB)**, Viyana, Avusturya, 770–781, 2007.

[67]. Xu, Y., Wang, K., Fu, A.W., Wong, R.C., "Publishing Skewed Sensitive Microdata", **Proceedings of the 2010 SIAM International Conference on Data Mining**, Columbus, Ohio, ABD, 84-93, 2010.

[68]. Wang, H., Han, J., Wang, J., Wang, L., "(l, e)-Diversity - A Privacy Preserving Model to Resist Semantic Similarity Attack", **Journal of Computers,** 59-64, 2014.

[69]. Wong, R.C, Fu, A.W, Wang, K., Pei, J., "Minimality attack in privacy preserving data publishing", **In Proceedings of the International Conference on Very Large Data Bases (VLDB)**, Viyana, Avusturya, 543–554, 2007.

[70]. Kifer, D., "Attacks on privacy and deFinetti's theorem", **In Proceedings of the 2009 ACM SIGMOD International Conference on Management of data**, Rhode Island, ABD, 127–138, 2009.

[71]. İnternet: Wikipedia, "de Finetti's theorem", http://en.wikipedia.org/wiki/De_Finetti%27s_theorem.

[72]. Sweeney, L., "Datafly: A system for providing anonymity in medical data". **Proceedings of the IFIP TC11 WG11.3 Eleventh International Conference on Database Securty XI: Status and Prospects**, Londra, İngiltere, 356–381, 1998.

[73]. Sweeney, L. "Achieving k-anonymity privacy protection using generalization and suppression", **International Journal of Uncertainty, Fuzziness and Knowledge-base Systems**, Cilt 10, Sayı 5, 571–588, 2002.

[74]. Wang, K., Fung, B. C. M., "Anonymizing sequential releases", **Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD)**, Philadelphia, ABD, 414–423, 2006.

[75]. Iyengar, V. S., "Transforming data to satisfy privacy constraints".**Proceedings of the 8th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)**, Edmonton, Alberta, Kanada, 279–288, 2002.

[76]. Bayardo, R. J., Agrawal, R., "Data privacy through optimal k-anonymization", **Proceedings of IEEE 21st International Conference on Data Enginering (ICDE 2005)**, Tokyo, Japonya, 217–228, 2005.

[77]. Fung, B. C. M., Wang, K., Yu, P. S., "Top-down specialization for information and privacy preservation". **Proceedings of IEEE 21st International Conference on Data Engineering (ICDE 2005)**, Tokyo, Japonya, 205–216, 2005.

[78]. Fung, B. C. M., Wang, K., Yu, P. S., "Anonymizing classification data for privacy preservation", **IEEE Transactions on Knowledge and Data Engineering,** Cilt 19, Sayı 5, 711–725, 2007.