Big data texnologiyalarında verilənlərin təhlükəsizliyinin əsas məsələləri

Aygül Fəxrəddinqızı
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
aygul.fexreddin@gmail.com

***Xülasə*— Verilənlər iqtisadiyyatın, sənayenin, təşkilatın, biznesin və fərdin vacib bir hissəsi olmaqla yanaşı İnternetin, ağıllı mobil qurğuların və sosial şəbəkələrin istifadəsi ilə günbəgün artır. Böyük verilənlər ölçüsü və mürəkkəbliyi ilə birlikdə iri həcmli verilənlər toplusudur və Verilənlər Bazasının İdarəetmə Sisiteminin (VBİS) saxlanması, idarə edilməsi və təhlil edilməsi olduqca çətindir. Böyük verilənlər küyün yığılması, ölçmə səhvləri, təhlükəsizlik və gizlilik də daxil olmaqla bir sıra problemləri özündə birləşdirir. Məqalənin əsas istiqaməti böyük verilənlərlə əlaqədar olan təhlükəsizlik və gizlilik məsələlərinə yönəldilmişdir. Məqalədə böyük ölçülü verilənlər konsepsiyası və bu verilənlərin təhlükəsizlik məsələləri nəzərdən keçirilir.**

Açar sözlər*— *Böyük ölçülü verilənlər; big datanın təhlükəsizliyi; verilənlərin gizliliyi.

I. GİRİŞ

Texnoloji inkişaf və İnternetin geniş yayılması, o cümlədən, mövcud məlumatların həcminin artması böyük ölçülü verilənlər konsepsiyasının meydana gəlməsinə səbəb olmuşdur. Böyük ölçülü verilənlər mənbələrdən əldə olunan iri həcmli, mürəkkəb verilənlər toplusudur. Bu məlumatlar toplusu o qədər böyükdür ki, onları artıq ənənəvi texnologiyalar vasitəsilə emal etmək mümkün deyil. Bu verilənlərin emalı üçün yeni texnologiyalardan istifadə etmək zərurəti yaranmışdır [1].

Böyük verilənlər konsepsiyasına ədəbiyyatda “Big Data” da deyirlər. Big datanın analizi vasitəsilə qiymətli məlumatlar və faydalı biliklər əldə etmək olar. Ancaq Big Datanın inkişafı təhlükəsizlik və gizlilik risklərini də özü ilə bərabər gətirir. Məsələn, online aldığımız məhsullar alış-veriş veb-saytı tərəfindən izlənilir; sosial şəbəkələrdə paylaşdığımız fikirlərimiz və şəxsi şəkillərimiz sayt tərəfindən qeyd olunur; şəxsi mülkiyyət haqqında məlumat bankların nəzarəti altında olur. Bu da öz növbəsində məlumatların toplanması, saxlanması və istifadəsi zamanı fərdi məlumatların asanlıqla sızmasına və məlumatların klassifikasiyası zamanı çətinliklərə səbəb olur. Aydındır ki, fərdi məlumatların sızması ciddi məsələdir. Böyük verilənlərin təhlükəsizliyini necə təmin etmək və gizliliyini qorumaq cari araşdırmalar mərhələsində ən aktual problemlərdən birinə çevrilmişdir. Bu verilənlərin təhlükəsizlik və gizlilik problemlərini necə həll etmək tədqiqatçıların əsas tədqiqat istiqamətinə çevrilmişdir [2]. Bunları nəzərə alaraq məqalədə II bölmədə Big Data təhlükəsizliyinin təmin edilməsi üçün mövcud yanaşmalar və metodlar, III bölmədə böyük ölçülü verilənlər konsepsiyası, IV bölmədə Big Data təhlükəsizliyində əsas məsələlərdən bəhs olunur.

II. ƏLAQƏLİ İŞLƏR

Bəzi tədqiqatçılar Big Datanın təhlükəsizliyini təmin etmək məqsədilə atributların seçilməsi üçün metod təklif etdilər. Metod təhlükəylə üzləşəcək dəyərli məlumatları çıxararaq, böyük məlumatları təhlil edir [3]. Əsasən, bir-birilə uyğunluq yaradan məlumatlar çıxarılır. Belə ki, metod məlumatların klassifikasiya problemini həll etmir, yəni şəkillər, sənədlər, cədvəllər və real vaxt məlumatları (məsələn, VoIP rabitə) kimi müxtəlif tipli məlumatların emalına baxılmır. [4,5]-də aparılan tədqiqatda Bulud şəbəkələrindən istifadə edilərək, böyük məlumatların qarşılaşdığı təhlükəsizlik problemləri araşdırılır. Tədqiqatçılar təklif etdikləri metodda şəbəkə təhlükəsizliyi, məlumat təhlükəsizliyi və gizlilik kimi məsələləri araşdıraraq, Bulud şəbəkələrində yayılmış Big Datanın tərkibinə daxil olmaq üçün yeni təhlükəsizlik modeli hazırlayıblar. Təklif olunan təhlükəsizlik modeli avtonom məlumatların qorunmasına yönəldilmiş və G-Hadoop paylanmış hesablama mühitində hazırlanmışdır. [6,7]-də aparılan tədqiqatda Storage-as-a-Service (STaaS) vasitəsilə paylanmış bulud anbarlarına müraciət edərək, Bulud şəbəkələri konsepsiyasında Big Datanın gizlilik məsələləri nəzərdən keçirilib. Həmçinin, Bulud Sistemləri və Əşyaların İnternetin (IoT) -də yaranan Big Data problemlərinə də toxunulub. Bundan əlavə, [8]-də gizlilik probleminə yönəldilmiş və Dynamic Data Encryption Strategiyası (D2ES) adlı verilənlərin şifrələnməsi üçün metod təklif edilmişdir. Metod vaxt məhdudiyyəti altında gizlilik təsnifatı metodlarından istifadə edərək məlumatları seçir və kodlaşdırır.

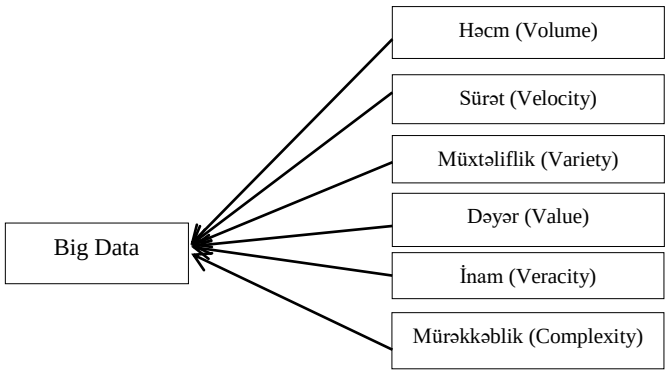
III. BÖYÜK ÖLÇÜLÜ VERİLƏNLƏR KONSEPSİYASI

Big data ənənəvi emal texnologiyalarından istifadə etməklə emalı çətin və ya qeyri-mümkün olan iri həcmli, sürətli və mürəkkəb verilənləri ifadə edir. Big datanın ən vacib əlamətlərindən biri verilənlərdən faydalı məlumat əldə etmək üçün yeni texnologiyaların istifadəsi və müxtəlif mənbələrdən verilənləri birləşdirmək qabiliyyətidir. Analitiklər üçün böyük miqdarda verilən əldə etmək və bu verilənləri saxlamaq məsələsi uzun müddətdir ki, aktualdır. Onu da qeyd edək ki, verilənlər ənənəvi olaraq relyativ verilənlər bazası kimi

DOI: 10.25045/NCInfoSec.2019.56

strukturlaşdırılmış formatda saxlanılır. Ancaq İnformasiya Kommunikasiya Texnologiyalarının (İKT) inkişafı ilə yeni bir tendensiya meydana gəlmişdir: mövcud verilənlərin həcmini strukturlaşdırılmamış və ya yarım strukturlaşdırılmış formatda saxlamaq [9].

Böyük ölçülü verilənləri qiymətləndirəndə onun müəyyən meyarları nəzərə alınmalıdır. Ədəbiyyatda Big Datanın meyarları dedikdə Doung Laneyin 6V xarakteristikası nəzərə çarpır (Şəkil 1). Bu xarakteristikaya bunlar daxildir [10]: Sürət, Həcm, Müxtəliflik, Dəyər, İnam, Mürəkkəblik



Şəkil 1. Big Data-nın 6V xarakteristikası

- ***Həcm.*** Həcm xarakteristikasına görə böyük verilənlərin analizində istifadə ediləcək alqoritmlərin seçimi zamanı, aşağıdakı kriteriyalar diqqətə alınmalıdır: verilənlərin ölçüsü, yüksək ölçü məsələsi və outlierların/kənaraçıxmaların emalı.
- ***Müxtəliflik.*** Ənənəvi VBİS ilə xüsusi struktura malik olmayan məlumatların saxlanılması və emalı çətin olsa da, “Big Data” mühitində bunun müxtəlif üsulları vardır. Müxtəliflik xarakteristikasına görə analiz üçün istifadə ediləcək alqoritmlərin seçimi zamanı, aşağıdakı kriteriyalar diqqətə alınmalıdır: verilənlərin tipi və klasterin forması.
- ***Sürət.*** Big Data prosesleri adətən saniyələr içərisində həyata keçirilir, yəni Big Data texnologiyaları sürətli formada verilənləri qəbul etməli, bu verilənlər üzərində görə biləcəyi işləri müəyyənləşdirib, analiz etməlidir. Əsasən Big Datada proseslər **real zaman anında**, **real zamana yaxın**, **batch** (müəyyən bloklar) şəklində aparılır. Sürət böyük həcmli verilənlərin analizində klasterləşdirmə alqoritminin sürətini ifadə edir və sürət xassəsinə uyğun klasterləşdirmə alqoritmini seçmək üçün aşağıdakı meyarlar nəzərə alınır: alqoritmin mürəkkəbliyi və işləmə müddəti.
- ***İnam.*** Böyük ölçülü verilənlər müxtəlif mənbələrdən əldə edilir, buna görə verilənlərin doğruluğunu / keyfiyyətini yoxlamaq tələb olunur.
- ***Qiymət.*** Giriş parametrləri klasterləşdirmə alqoritmlərinin verilənləri dəqiq yerinə yetirmək və daha az hesablama aparmaqla klaster əmələ gətirməsi mühüm rol oynayır.
- ***Mürəkkəblik.*** Müxtəlif mənbələrdən alınan verilənlər fərqli quruluşa malikdir və həmin verilənlər bir-birilə əlaqələndirilməlidir. Eyni verilənlər bazasında və ya çoxlu verilənlər bazasında müxtəlif dəyişənlər arasında yaranan əlaqəyə görə mürəkkəb verilənləri emal etmək və təhlil etmək çətindir. Ənənəvi analiz alətləri dəyişənlər arasındakı münasibətləri aydınlaşdırmaq üçün çoxlu sayda iterasiya tələb edir. Dəyişənlər arasındakı bu əlaqəni tapmaq üçün mürəkkəb verilənlərin təhlili Big Data üçün üstünlüklər gətirir.

IV. BİG DATA TƏHLÜKƏSİZLİYİNDƏ ƏSAS MƏSƏLƏLƏR

İnternet, smart qurğular və sosial şəbəkələrdən istifadə nəticəsində verilənlərin sayı gündən-günə artır. Bu verilənlər Big Datanın həcmi və ya müxtəlifliyinə təsir etsə də, həmçinin verilənlərin təhlükəsizliyi və gizliliyi ilə bağlı məsələləri də ortaya qoyur. Böyük ölçülü verilənlər əhəmiyyətli və mürəkkəb bir mövzu olduğundan, təhlükəsizlik və gizlilik məsələlərinin ortaya çıxması demək olar ki, təbiidir. Böyük ölçülü verilənlərin təhlükəsizliyini və gizliliyini təmin etmək ən vacib problemlərdən biridir. Bu problemlər verilən, prosses, idarəetmə, ötürmə və saxlama ilə bağlıdır.

“Cloud Security Alliance” təşkilatındakı Big Data İş Qrupuna görə, Big Data təhlükəsizliyinin prinsipcə üç fərqli tərəfi var: infrastruktur təhlükəsizliyi, verilənlərin gizliliyi, verilənlərin idarə edilməsi [11].

İnfrastruktur təhlükəsizliyi – Burada ən çox istifadə olunan Hadoop texnologiyasına əsaslanan arxitektura xüsusilə vurğulanır.

Verilənlərin gizliliyi – həm istifadəçilər həm də təşkilatlar üçün ən böyük narahatlıqlar gətirəcək məsələlərdən biridir. Onlayn mühitdə çoxlu sayda fərdi məlumatlar toplana bilər. Bu məlumatlardan isə bədniyyətli təşkilatlar (insanlar) öz mənafeyləri üçün istifadə edə bilərlər. Bu problemi aradan qaldırmaq üçün bir sıra metodlar işlənmişdir. Bu problemin həll olunmasının əsas yollarını göstərən bir neçə məsələ vardır: kriptoqrafiya, konfidensiallıq,gizliliyi qoruyan sorğular, anonimizasiya, sosial şəbəkələrdə gizlilik, diferensial gizlilik [12] və s. Bunlardan bəziləri ilə tanış olaq:

Kriptoqrafiya Big Data sistemində məlumatların gizliliyinin təmin edilməsi məsələsində ən çox istifadə olunan həll yoludur. Bu üsulda **bitmap** şifrələmə sxemi vasitəsilə verilənləri uzun müddət qorumaq olur. Sosial şəbəkələr hazırda böyük populyarlıq qazanmışdır və İnternetdən istifadə edən hər kəsin ən azı bir sosial şəbəkə hesabı vardır. İstifadəçilər sosial şəbəkələrdə fərdi məlumatlarını, şəxsi düşüncələrini paylaşırlar. Bu məlumatlar şəxsi məxfiliyimiz üçün də risklər yaradır [13]. Bu problemi həll etmək asan məsələ olmasa da bəzi tədqiqatçılar yeni konsepsiyalar təklif edərək verilənlərin məxfiliyini qismən təmin edə bilirlər.

Diferensial gizlilik ilk dəfə 2006-cı ildə Dwork, Nissim tərəfindən hazırlanmış hücum vasitəsidir. Əvvəlcə onu qeyd edək ki, diferensial gizlilik verilənlər bazasına deyil, sorğulara tətbiq olunur. Bu metodunun əsas üstünlüyü **küy** üsulundan istifadə edilməsidir. Nəzərə alınacaq bir məsələ var ki, diferensial gizliliyi qorumaq üçün sorğular nə qədər geniş olarsa, o qədər küy tətbiq olunmalıdır. Bu da istifadəçilərin

şəxsiyyətlərini müəyyənləşdirmək qabiliyyətini minimuma endirir və bir sıra verilənlərin analizini asanlaşdırır.

Diferensial gizliliyin tərifinə nəzər salsaq, burada iki çoxluğa baxılır. Bu çoxluqlar D və D' kimi işarə olunur. Bunlardan birində fərdin məlumatları yer alıb, digərdində isə həmin məlumatlar silinib və ya fərqli məlumatlarla əvəz olunub. Məsələnin əsas məqsədi prosesin sonunda D və D'-də olan fərdi məlumatlar arasında oxşarlıq maksimum olsun. Oxşarlıq ε parametri vasitəsilə ölçülür. ε parametri D və D'-dən alınmış nəticələrin paylanmasının nə dərəcədə yaxın olmasını ölçür və ε parametri kiçik olduqda gizliliyin təhlükəsizlik səviyyəsi yüksək olur [14].

Konfidensiallıq gizlilik termini ilə eyni olsa da, bunlar hüquqi baxımdan tamamilə fərqlənirlər. Konfidensiallıq müvəkkil, həkim, terapevt, vəkil və ya müştərinin razılığı olmadan fərdi məlumatların səlahiyyəti olmayan üçüncü şəxslərlə paylaşılmamasıdır. Məsələn, onlayn ödəmə hesabınızla bağlı məlumatların başqa şəxsin əlinə keçməsi informasiyanın konfidensiallığının pozulmasıdır.

Digər tərəfdən, gizlilik isə fərdi məsələlərə və fərdi məlumatlara müdaxilə etmək hüququna aiddir. Konfidensiallıq etik cəhətdən doğru olsa da, gizlilik qanunlarda yer alan bir hüquqdur.

Verilənlərin idarə edilməsi – təşkilat və ya müəssisə tərəfindən yaradılan verilənlərin alınması, toplanması, təşkili və saxlanması prosesidir [15]. Effektiv verilənlərin idarə edilməsi müəssisə tətbiqetmələrini idarə edən və təşkilat idarəçiləri, iş menecerləri və istifadəçilər tərəfindən operativ qərarvermə və strateji planlaşdırmanı idarə etmək üçün analitik məlumat təqdim edən İnformasiya Texnologiyaları (İT) sistemlərinin istifadəsində vacib rol oynayır.

Ənənəvi emal texnologiyaları Big Data konsepsiyasının təhlükəsizliklə bağlı gətirdiyi problemlər qarşısında çarəsizdir. şifrləmə metodları, girişlərin idarə edilməsi, firewalllar, ötürmə mühitlərinin təhlükəsizliyi və s. [16] kimi isitfadə edilən metodlara müdaxilələr edilə bilər. Buna görə də bir çox aspektdən Big Data proseslərinə nəzarət etmək və təhlükəsizliyini təmin etmək məqsədilə yeni üsul və vaitələrin işlənməsinə ehtiyac vardır.

NƏTİCƏ

Big Data analitikasının əsas məqsədi çox sayda xam məlumatdan faydalı məlumat əldə etməkdir. Lakin bu zaman Big Data ilə bağlı bir sıra təhlükəsizlik məsələləri meydana çıxır. Məqalədə Big Data təhlükəsizliyi ilə bağlı əsas məsələlər nəzərdən keçirilmişdir. Burada şübhəli davranışların tez bir zamanda aşkarlanması üçün şəbəkə trafikinin daim izlənməsinin vacib olduğu vurğulanmış və bütün rabitənin təhlükəsiz kanallar üzərində aparılması, verilənlər bazasının yayımlanmadan əvvəl fərdi məlumatların “maskalanması” tövsiyyə olunmuşdur. Böyük verilənlərin məxfiliyi və təhlükəsizliyi istiqamətində daha çox tədqiqatlar aparılmalıdır. Gələcək işlərimizdə bu istiqamətdə, yəni verilənlərin təhlükəsizliyini təmin etmək məqsədilə yeni yanaşmaların verilməsi və klasterləşmə alqoritmlərinin müqayisəli təhlili nəzərdə tutulmuşdur.

İSTİNADLAR

[1] Z. Dongpo, “Big data security and privacy protection,” In 8th International Conference on Management and Computer Science , vol.15, no. 2, 2018, pp.275–278.
[2] F.Yan, “Big Data Security and Privacy Protection,”. Electronic Technology and Software Engineering, 2016, pp. 227-235.
[3] I. Narasimha, A. Sailaja, and S. Ravuri, “Security Issues Associated with Big Data in Cloud Computing,” International Journal of Network Security and Its Applications, vol. 6, no. 3, 2014, pp. 45–56.
[4] S.-H. Kim, N.-U. Kim, and T.-M. Chung, “Attribute relationship evaluation methodology for big data security,” in Proceedings of the 2013 3rd International Conference on IT Convergence and Security, ICITCS 2013, China, December 2013.
[5] J. Zhao, L. Wang, J. Tao et al., “A security framework in G-Hadoop for big data computing across distributed cloud data centres,” Journal of Computer and System Sciences, vol. 80, no. 5, 2014, pp. 994–1007.
[6] K. Gai, M. Qiu, and H. Zhao, “Security-Aware Efficient Mass Distributed Storage Approach for Cloud Systems in Big Data,” in 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), New York, NY, Usa, April 2016, pp. 140–145
[7] Liu, C. Yang, X. Zhang, and J. Chen, “External integrity verification for outsourced big data in cloud and IoT: a big picture,” Future Generation Computer Systems, vol. 49, 2015, pp. 58–67.
[8] K. Gai, M. Qiu, H. Zhao, and J. Xiong, “Privacy-Aware Adaptive Data Encryption Strategy of Big Data in Cloud Computing,” in Proceedings of the 3rd IEEE International Conference on Cyber Security, China, June 2016,pp. 273–278.
[9] R.Alıquliyev , G.Niftəliyeva, “E-Dövlətin Big Data Mənbələri,” “Big data: imkanları, multidissiplinar problemləri və perspektivləri” I respublika elmi-praktiki konfransı Bakı şəhəri, 2016, pp. 78-80.
[10] R.Patel,“Security & privacy of Big data: Big challenge for an organization,” IJSRD - International Journal for Scientific Research & Development, vol. 3, no. 11, 2016, pp.303-305.
[11] A.A. Cardenas, P.K. Manadhata, S.P. Rajan, “Big Data Analytics for Security”, IEEE Security & Privacy, vol. 11, no. 6, 2013, pp. 74 – 76.
[12] J.Moreno, M. A. Serrano and E. Fernández-Medina, “Main Issues in Big Data Security,” Department of Computer Science & Engineering, vol.4, 2016, pp.265-279.
[13] B.B.Rada, N.Akbarzadeh, P. A. and Y.Khakbiz, “Security and Privacy Challenges in Big Data Era,” International Journal of Control Theory and Applications, vol. 9, no.43, 2016, pp.437-448.
[14] Y.İmamverdiyev, “Big data və fərdi məlumatların təhlükəsizliyi,” “Big data: imkanları, multidissiplinar problemləri və perspektivləri” I respublika elmi-praktiki konfransı Bakı şəhəri, 2016, pp. 109-113.
[15] B. Matturdi, X. Zhou, S. Li, F. Lin, “Big Data security and privacy: A review,” Big Data, Cloud & Mobile Computing, China Communications vol.11, no.14, 2014, pp. 135 – 145.
[16] H. Cheng, C. Rong, K. Hwang, W. Wang, Y. Li, “Secure big data storage and sharing scheme for cloud tenants,” Communications, China, vol. 12, no.6, 2015, pp. 106 – 115.

MAIN SECURITY ISSUES IN BIG DATA TECHNOLOGIES

Aygul Fakhraddingizi

Institute of Information Technology of ANAS, Baku, Azerbaijan

aygul.fexreddin@gmail.com

Abstract – Data is becoming increasingly important as a significant part of the economy, industry, organization, business, and individual, with the use of the Internet, smart devices and social networks. This is a big data collection with a large volume and complexity of data, and it is extremely difficult to maintain, manage and analyze a Database Management System (DBMS). Big data contains a number of problems, including noise accumulation, measurement errors, security and privacy. The focus of the article is on security and privacy related to big data.

Keywords – *big data; big data security; data privacy.*