UOT 004.056

Hacırəhimova M.Ş.
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
makrufa@science.az

"BIG DATA"TEXNOLOGİYALARI VƏ İNFORMASİYA TƏHLÜKƏSİZLİYİ PROBLEMLƏRİ

İnformasiyanın emalında yeni eranı əks etdirən"Big Data"mövzusu biznes mühitində, kütləvi informasiya vasitələrində, həmçinin elmi ədəbiyyatda geniş müzakirə olunmaqdadır. "Big data" idarəçilik, səhiyyə, elm, biznes və kommersiya, sənaye və digər sahələrdə inqilabi dəyişikliklər edə biləcək bir texnologiyadır. Bu texnologiya bir tərəfdən cəmiyyət üçün yeni imkanlar açır, digər tərəfdən isə yeni təhlükəsizlik problemləri yaradır. Məqalədə"big data" texnologiyalarının qısa xülasəsi verilir. "Big data" analitikanın faydaları, bəzi təhlükəsizlik problemləri tədqiq olunur. Fərdi məlumatların qorunması baxımından onun yaratdığı yeni etik problemlərə baxılır və bəzi tövsiyələr verilir.

***Açar sözlər**: big data, big data analitikası, təhlükəsizlik, anonimlik, fərdi məlumatlar, identifikasiya, de-identifikasiya, şifrləmə.*

Giriş

Böyük verilənlər (BV) problemi yeni problem deyil. Texniki və texnoloji inkişaf ilə əlaqədar olaraq verilənlərin sürətlə artması və onların emalı problemləri hələ keçən əsrin 40-cı illərinə təsadüf edir. Şübhəsiz, o vaxt ilə müqayisədə, XXI əsrin əvvəllərindən başlayaraq rəqəmsal verilənlər hər il həndəsi silsilə ilə artmaqdadır [1]. Bunu IDC, Gartner və s. kimi analitik şirkətlərin hesabatları da təsdiq edir [2]. Veb, sosial şəbəkələr, mobil qurğular, kredit kartları vasitəsilə edilən tranzaksiyalar və s. rəqəmsal verilənlər axınının artmasına gətirib çıxarmış, informasiya bolluğu yaranmış, dünya sanki informasiya ilə doldurulmuşdur. Belə ki, bəşəriyyətin mövcudluğundan 2003-cü ilə qədər olan dövrdə dünyada cəmi 5 ekzabayt ($5x10^9$*Qbayt*) məlumat generasiya olunmuşdursa, 2015-ci ildə dünyada informasiyanın həcminin 8,1 zetabayt ($8,1x10^3$*ekzabayt*) olacağı, növbəti hər il 40% artaraq 2020-ci ildə 44 zetabayta çatacağı proqnozlaşdırılır [2]. Nəticə etibarı ilə verilənlərin emalı, saxlanılması və istifadəsində yeni eranı əks etdirən "böyük verilənlər" (*ing. big data*) termini meydana çıxmışdır [3]. Bu termin dövlət qurumları, böyük şirkətlər və s. tərəfindən yaradılan və sonradan müxtəlif məqsədlər üçün analiz edilən çox nəhəng rəqəmsal verilənlər toplusuna aid edilir ki, onları ənənəvi verilənlər bazası və alətlərin köməyi ilə emal etmək mümkün olmur. Bu verilənlərin emalı üçün kifayət qədər saxlama tutumu və hesablama gücü tələb olunur. Problemin həlli məqsədi ilə informasiya texnologiyaları sənayesinin nəhənglərindən olan Google şirkəti tərəfindən 2004-cü ildə Google File System və MapReduce [4] proqram-aparat platforması yaradılmışdır. Bu platforma əsasında paylanmış hesablama mühitində BV-in emalı və analizi üçün açıq kodlu Apache Hadoop və Hadoop File System [5] proqram platformaları işlənmiş və bununla da "big data" texnologiyalarının əsası qoyulmuşdur.

İstənilən texnoloji nailiyyət rifah və ya bəd niyyətlər üçün istifadə oluna bilər. "Big data"nın da digər texnologiyalar kimi iki tərəfi: faydaları və təhlükələri vardır. BV bir tərəfdən cəmiyyətin bütün sahələrini kökündən dəyişə biləcək təsirə malik bilik mənbəyidir, biznesə yeni üfüqlər açır. Digər tərəfdən, informasiya nə qədər rəqəmsallaşdırılırsa və əlavə informasiya toplanırsa, bir o qədər əlçatan olur və onun istifadəçilərinin sayı da çoxalır, bu isə bədniyyətlilər üçün potensial imkanlar yaradır. İnformasiyanın təhlükəsizliyi baxımından bədniyyətlilər tərəfindən informasiyanın oğurlanması, təhrif olunması və şəbəkələrin sındırılması, fərdi məlumatların asanlıqla ələ keçməsi və s. kimi təhlükələr yaradır [6]. İnsanların razılığı olmadan onlara məxsus fərdi məlumatlar analiz olunur. Bu da öz növbəsində etik və hüquqi cəhətdən yolverilməzdir, təhlükəsizlik və gizlilik baxımından çox ciddi problemdir. Problemlərə müxtəlif

prizmadan yanaşmaq olar: informasiya təhlükəsizliyi üçün "big data" analitikanın tətbiqi və ya "big data" analitikada informasiya təhlükəsizliyi [7–11]. Bu məsələlər isə elmi tədqiqatçılar qarşısında duran ən akual məsələlərdəndir. Təqdim olunan məqalədə də məqsəd "big data" texnologiyalarının əsas təhlükəsizlik problemlərini və gizli məlumatların qorunmasındakı problemləri analiz etməkdir.

"Big data" analitika və informasiya təhlükəsizliyi problemləri

Bu gün "big data" adlandırılan informasiya bolluğu həqiqətən də mövcuddur. "Big data"nı təyin etməyə və digər verilənlərdən fərqləndirməyə kömək edən ilk model «3V»lər adlanır [12]. Bu model çox böyük sürətlə (*velocity*) və müxtəlif mənbələrdən (*variety*) toplanan çox böyük həcmdə (*volume*) verilənləri daha səmərəli istifadə etmək, saxlamaq, analiz edərək ondan daha qiymətli informasiyanı əldə etmək ideyasını özündə əks etdirir. Qeyd etmək lazımdır ki, IBM şirkəti verilənlərin həqiqiliyini əsas götürərək, 4-cü "v" (veracity), Oracle isə BV-nin dəyərini qeyd etməklə, 5-ci "v"-ni (value) daxil etmişdir [1, 2, 13]. Son zamanlar mütəxəssislər tərəfindən "v"-lərin sayı artırılmaqdadır.

BV-nin əsas ideyası çox böyük həcmdə verilənləri real-vaxt ərzində emal və analiz etmək imkanının olmasıdır. İnformasiyanın həcminin artması insanlarda məyusluq yaratmamalıdır. Əksinə, ona təbii xammal, resurs kimi baxmaq lazımdır. Qeyd edildiyi kimi, BV dünyanı dəyişə biləcək gücə, biznesdə, dövlət idarəçiliyində inqilabi yeniliklərə qadirdir. Çünki bu xam verilənlərdə elmi kəşflərə səbəb ola biləcək dərin bilik toplanmışdır. Ancaq, bu resursdan maksimum istifadə etməklə, cəmiyyət və biznes sahəsində dəyər yaratmaq üçün yeni nəsil analitik texnologiyalara ehtiyac vardır. Bu baxımdan, BV mövzusu istər biznes sektorunda, istər kütləvi informasiya vasitələrində, istərsə də dövlət qurumlarında qərar qəbul edən şəxslər və siyasətçilər tərəfindən böyük diqqət çəkir. BV-nin mövcud olması ilə o, elmi mühitdə yeni keyfiyyətdə tədqiqat sahəsi kimi izlənməyə başlanmışdır. Real vaxta maksimum yaxın rejimdə verilənlərin analizinə olan tələbat müxtəlif parametrlər, xarakteristikalar, hadisələr və s. arasındakı korrelyasiyanı tapmağa, klassifikasiya və analitik hesabatlar və bunun əsasında proqnozların verilməsinə imkan verən BV analitikasının (*Big Data Analytics*) yaranmasına gətirib çıxardı [1, 13–16]. Qeyd etmək lazımdır ki, BV-nin analitikası nəticəsində səhv korrelyasiyalar da ola bilər.

BV analitikası korporativ maraqlar baxımından biznes-proseslərin səmərəliliyinin artırılmasına, marketinq işlərinin yaxşılaşdırılmasına imkan verir. Müəssisələrdə BV-nin toplanması və analizinin köməyi ilə gəlirləri və xərcləri optimal idarə etmək, maliyyə göstəricilərini yaxşılaşdırmaq və şəffaflığı yüksəltmək mümkündür. "*Machine-to-machine"(M2M*) kimi ikitərəfli qarşılıqlı əlaqə nəticəsində müxtəlif mənbələrdən və müxtəlif formatda (*strukturlaşdırılmış və strukturlaşdırılmamış*) fasiləsiz olaraq generasiya olunan verilənlərin birgə analizi və onlardan yeni biliklərin və faydalı məlumatların əldə olunması yeni elmi kəşflərin edilməsində son dərəcə əhəmiyyətlidir. Dövlət və özəl təşkilatlarda əsaslandırılmış düzgün qərarların qəbul edilməsində, hüquq qaydalarının qorunmasında, milli təhlükəsizlik, terrorizm faktlarının aşkar edilməsində, xəstəlik epidemiyalarının əvvəlcədən müəyyən edilməsində, insanların gizli davranışlarını üzə çıxarmaq, məqsəd və niyyətlərini anlamaq, onların digər insanlarla, ətraf mühitlə qarşılıqlı əlaqəsini başa düşməkdə, maliyyə s**e**ktorunda milli səviyyədə iqtisadi riskləri daha yaxşı anlamaq, siyasətçiləri və tənzimləyici orqanları istiqamətləndirmək və risk sistemlərini daha yaxşı idarə etməkdə bu texnologiyadan istifadə önəmlidir [1].

Gartner analitik şirkətinin tədqiqatlarında qeyd edildiyi kimi, BV-nin analizi cinayətlərin və təhlükəsizliyin pozulması hallarını üzə çıxarmaqda əsas rol oynayacaq, 2016-cı ilə kimi böyük şirkətlərin 25%-i kiber-təhlükəsizlik sistemlərində "big data" analitikanı istifadə edəcəklər. Gartner ekspertləri hesab edirlər ki, "Big Data" analitikası informasiya təhlükəsizliyi sistemlərini daha etibarlı etməyə imkan verəcəkdir [17].

Ancaq, "big data" texnologiyalarının tətbiqi mövcud təhlükəsizlik modellərinin

köhnəldiyini üzə çıxarmışdır. 15 il əvvəl istifadə olunan təhlükəsizlik yanaşması bu gün üçün adekvat deyildir. Eyni zamanda, BV-nin həcm, müxtəliflik və sürət kimi xarakteristikaları isə təhlükəsizlik və gizlilik problemini daha da kəskinləşdirir [18, 19]. Verilənlərin mənbələrinin müxtəlifliyi və axın şəklində toplanması, geniş miqyaslı "bulud hesablamaları" infrastrukturu və böyük həcmdə informasiyaların "buludlara" miqrasiyası, təhlükəsizlik sistemlərinnin "zəif" cəhətlərini üzə çıxarmışdır. Belə ki, BV-nin genişlənməsi ilə ənənəvi təhlükəsizlik mexanizmləri kifayət deyildir. Çünki verilənlərin axını çox çevik və sürətli təhlükəsizlik həlləri tələb edir.

Bulud Təhlükəsizlik Alyansı (CSA – *Cloud Security Alliance)* tərəfindən "Big data" sistemlərinin on təhlükəsizlik problemi verilmiş və onlar dörd qrup üzrə təsnifatlandırılmışdır [20]:

- infrastrukturun təhlükəsizliyi (paylanmış proqram mühitinin təhlükəsizlik tədbirləri, qeyri-relyasiya tipli verilənlər xəzinəsi üçün ən yaxşı təhlükəsizlik təcrübəsi);
- verilənlərin gizliliyi (gizliliyi saxlayan miqyaslanan və kompozit data mining və analitika, kriptoqrafiya ilə həyata keçirilən girişə nəzarət və təhlükəsiz kommunikasiya, girişə qranulyar nəzarət);
- verilənlərin idarə edilməsi (verilənlərin saxlanması və ötürülməsi jurnalına nəzarəti möhkəmləndirmək, verilənlərin mənşəyi, qranulyar auditlər);
- tamlıq və reaktiv təhlükəsizlik (real-vaxt rejimində təhlükəsizliyin monitorinqi, girişin yoxlanılması/ filtirlənməsi).

Təsnifatdan da göründüyü kimi, "big data" sistemlərinin təhlükəsizlik infrastrukturunu təmin etmək üçün paylanmış hesablamalar və verilənlər xəzinəsi mühafizə olunmalıdır. Hər şeydən əvvəl, saxlanılan və gizli informasiyanın özünün təhlükəsizliyi üçün kriptoqrafiya və qranular giriş nəzarət vasitələrindən istifadə olunmalıdır. Böyük həcmli informasiyanın idarə olunması verilənlər xəzinəsi üçün verilənlərə effektiv nəzarət və mənşəyini müəyyən etmək üçün miqyaslanan və paylanmış həllər tələb edir. Nəhayət, müxtəlif nöqtələrdən axın şəklində daxil olan verilənlərin tamlığı yoxlanmalı və real-vaxt rejimində təhlükəsizlik insidentləri üzrə analitik təhlil aparılmalıdır.

Təhlükəsizlik və gizlilik problemlərinin həllində, adətən, üç məsələnin yerinə yetirilməsi lazımdır:

1. *Modelləşdirmə*: kiber-hücum və ya verilənlərin sızmaları ssenarilərini əhatə edən təhlükəsizlik modelini formalaşdırmaq;

2. *Analiz*: təhlükəsizlik modeli əsasında mümkün həllər tapmaq;

3. *Realizasiya*: tapılmış həlli mövcud infrastrukturda tətbiq etmək.

Təhlükəsizliyi təmin etmək üçün bütün Hadoop məhsullarında dörd səviyyəli təhlükəsizlik (*4-Layer Security*) modeli tətbiq edilir [21] (şək.1).

Təhlükəsizlik perimetri (*Perimeter Security*): istifadəçinin autentifikasiyasına cavab verir və şəbəkə təhlükəsizliyi problemini Massaçusets Texnologiya İnstitutu tərəfindən yaradılmış Kerberos şəbəkə protokolu vasitəsi ilə yerinə yetirir. Kerberos klient-server tətbiqlərində məxfi açarlı şifrləmə (*secret-key cryptography*) vasitəsi ilə etibarlı autentifikasiyanı təmin etmək üçün nəzərdə tutulmuşdur. Yəni, informasiya sistemlərinin təhlükəsizliyini təmin etmək üçün şəbəkə üzrə autentifikasiya və dayanıqlı kriptoqrafiya alətləri təqdim edir. Həqiqiliyin yoxlanmasında de-fakto standart hesab olunan Kerberos hər kəs üçün açıqdır – əlyetərlidir [10].

Verilənlərin əlyetərliyi (*Data Access Security*): istifadəçilərin ancaq verilənlərə giriş icazəsinin olduğunu, xidmət və resurslara isə olmadığını təmin etmək üçün nəzərdə tutulur.

Hesabatlılıq *(Accountability):* bu təhlükəsizlik səviyyəsinin ümumi məqsədi hesabatlılığı təşviq etməkdir. Hadoop-da administratorlara verilənlərə girişi audit etməyə imkan verir. Əlavə olaraq, verilənlərin mənşəyini, yəni verilənlərin hansı mənbədən daxil olduğunu da müəyyən etməyə imkan verir. Bu səviyyənin təhlükəsizliyini dəstəkləmək üçün Navigator adlanan xüsusi məhsul işlənib hazırlanmışdır.

Şəkil 1. Hadoop sistemində verilənlərin təhlükəsizlik modeli

Verilənlərin qorunması *(Data Protection)* sonuncu təhlükəsizlik aspektidir və verilənlərin şifrlənməsi, maskalanması və daha çox sahəni əhatə edir [8].

"Big data" və fərdi məlumatların qorunması

Təxminən 30 il bundan əvvəl bəlkə də insanların şəxsi həyatını mühafizə etmək və anonimliyini təmin etmək nisbətən asan idi. Çünki şəxsi informasiyaları toplayan və saxlayan avtomatlaşdırılmış sistemlər az idi, İnternet çox primitiv idi, hətta onun haqqında bilənlərin sayı belə az idi. Son illər isə hər şey dəyişmiş, rəqəmsallaşma *(GPS siqnalları, mobil telefonlar, e-mail, elektron alqı-satqı, sosial şəbəkələrdəki yazışmalar, elektron-tibb yazıları və s.)* geniş miqyas almışdır. Marketinq, bank, sığorta, tibb və s. sahələrdə müştərilər haqqında toplanan informasiya fərdi məlumatların gizliliyi və anonimliyini keçmişdə qoymuşdur.

"Big data" müstəsna faydalı informasiya versə də, təhlükəsizlik və gizlilik ilə bağlı yeni etik problemlər yaradır. "Data Mining" texnologiyaları inkişaf etdikcə və geniş tətbiq olunduqca, BV-dən daha dərin biliklərin əldə olunması son dərəcə həssas fərdi məlumatların qorunmasını təhlükə altında qoyur. İstər şirkətlər, istərsə də dövlət təşkilatları tərəfindən insanlar haqqında toplanan informasiya artıq nəzarətdən çıxmış, açıq və gizli verilənlər arasındakı sərhəd silinməkdədir. Bu da insanların şəxsi həyatlarının toxunulmazlığı baxımından ciddi problemdir [7].

Fərdi məlumatlar özləri açıq və konfidensial kateqoriyalara bölünür. Açıq fərdi məlumatlar kateqoriyasına müəyyən olunmuş qaydada adsızlaşdırılmış, subyekt tərəfindən açıq elan olunmuş və ya ümumi istifadə üçün yaradılmış informasiya sisteminə subyektin razılığı ilə onun barəsində daxil edilmiş məlumatlar aiddir. Şəxsin adı, soyadı və atasının adı daimi açıq fərdi məlumatdır. Konfidensial məlumatlardan fərqli olaraq, açıq kateqoriyalı fərdi məlumatların gizliliyinin təmin

edilməsi tələb olunmur. Konfidensial fərdi məlumatlar qanunla müəyyən olunmuş hallar istisna olmaqla, üçüncü şəxslərə yalnız subyektin razılığı əsasında verilə bilər [22].

Verilənlər bilavasitə insan fəaliyyətini əks etdirir. İnsanlar müalicə alır, elektron xidmətlərdən istifadə edir, veb saytlarda axtarışlar aparır, telefon zəngləri edirlər. Onların hansı coğrafi məkanda olduqları, ailə üzvləri ilə əlaqələri, siyasi fəaliyyətləri, sosial dairələri və maraqları öz smart telefonları və istifadə etdikləri proqram əlavələri ilə daim izlənilir və şirkətlər tərəfindən toplanır. Bu gün, bu məlumatların böyük hissəsi insanların razılığı olmadan toplanır və istifadə edilir. Sığorta şirkətləri və banklar müştərilərinin borcu olduğunu öyrənməklə onlara kredit verməkdən imtina edir, marketlər valideynlərdən qabaq qızın hamilə olduğunu müəyyən edirlər, həkim öz pasiyentinin elektron-tibbi məlumatları əsasında, hətta xəstənin gizlətmək istədiyi informasiyaya malik olur [23]. Bu da fərdi məlumatların qorunmasının prinsip və qaydaları ilə ziddiyyət təşkil edir.

Yarana biləcək təhlükələrdən biri də odur ki, insanlar təsadüf nəticəsində cinayətkar kimi məsuliyyətə cəlb oluna bilərlər. Belə ki, sosial şəbəkələrə onlayn proqramlar vasitəsilə yüklənən böyük həcmdə multimedia (*audio,video*) verilənlərinin yayılması səbəbindən istənilən insident zamanı təsadüfən kadra düşən, yoldan keçənlərin hüquqları pozula bilər. Məsələn, Bostonda partlayışların tədqiqatı zamanı, terror aktı yerindəki fotoşəkillər sosial şəbəkə saytlarında yerləşdirildiyi üçün bir neçə nəfər şübhəlilər sırasına düşmüşdü.

Bu gün müasir informasiya texnologiyaları faktorları: BV, analitika və bulud texnologiyalarını bir-birindən ayrı təsəvvür etmək mümkün deyildir. Bulud texnologiyaları verilənlərin saxlanmasında və böyük hesablamaların aparılmasında son dərəcə müvəffəqiyyətli yanaşmalardan hesab olunur [24]. Ancaq, hazırda verilənlərin buludda təhlükəsiz və effektiv saxlanması bulud hesablamaları sahəsində ən böyük problemdir. Buludda saxlanan verilənlərin etibarlı şəkildə qorunmasına zəmanət yoxdur. [25]-də bulud hesablama sistemində təhlükəsizliyi təmin edən etibarlı hesablama mühiti yaratmaq üçün metod təklif edilmişdir. Bu metod istifadəçilərə verilənlərini buludda təhlükəsiz və effektiv saxlamağa imkan verir. BV-nin idarə edilməsi və təhlükəsizlik problemləri verilənlərin ilkin olaraq buluda yükləndiyi zaman şifrlənmə və sıxma (compression) metodu ilə həll edilir. Kodlaşdırma texnologiyalarının tədqiqi sistemin etibarlılığını qoruyur, həm də təhlükəsizliyi və verilənlərin saxlama sistemlərindən istifadəni yaxşılaşdırır [25].

Lakin gizlilik problemi insanlara şəxsi məlumatlarını, ümumiyyətlə, informasiyanı buludlarda saxlamağa mane olmaqda davam edir. Bu problem BV-nin intellektual analizi və analitikasının (*big data mining and analytics*) inkişafı ilə daha da ciddiləşmişdir [16]. Çünki, burada relevant nəticə almaq üçün fərdiləşdirilmiş və lokal-bazaya əsaslanan xidmətlər kimi şəxsi informasiya tələb olunur. Şəxsə aid informasiya çox dəqiq yoxlamalara və profilləşmə narahatlığına, oğurlanma və itmə kimi risklərə məruz qalır [18].

Şəxsi verilənlərin təhlükəsizliyinə cavabdeh təşkilatlar, adətən, de-identifikasiya üsullarından, o cümlədən, anonimlik (*anonymization*), təxəllüs (*pseudonymization*), şifrləmə (*encryption*), açar-kodlaşdırma (*key-coding*) və s. istifadə edirlər. Anonimlik ad, ünvan və sosial sığorta (*müdafiə*) nömrələrini silməklə gizliliyi təmin edirsə, təxəllüs bu informasiyanı ləqəb və süni identifikasiya ilə əvəz edir. Açar-kodlaşdırma şəxsi informasiyanı kodlaşdırır və onun dekodlaşdırılması üçün açar yaradır. İnsanların şəxsi verilənlərini "Big Data" sistemlərinə təqdim etməsi istəyinin müəyyən məqamları vardır. Bunun üçün aşağıdakı təhlükəsizlik həlləri təmin olunmalıdır:

- "Big Data"-nın təhlükəsiz və gizliliyin qorunması ilə toplanması və emalı;
- "Big Data" analizinin təhlükəsiz mühitdə və gizliliyin qorunduğu şəkildə həyata keçirilməsi;
- "Big Data" sistemləri üçün verilənlərin saxlanılması və saxlanma siyasətinin təhlükəsiz (*və gizlilik rejimində*) tətbiqi.

Əks təqdirdə istifadəçilər verilənlərini "Big Data" sistemlərinə təqdim etməkdə tərəddüd edirlər. Konfidensial fərdi məlumatlar qanunvericilikdə nəzərdə tutulmuş tələblərə uyğun səviyyədə mülkiyyətçi, operator və bu məlumatlara giriş hüququ olan istifadəçilər tərəfindən mühafizə olunmalıdır. Fərdi məlumatların toplanılması, işlənilməsi və mühafizəsi hüquqi aktlarla tənzimlənməlidir [22, 26].

"Big data" analitika ayrı-ayrı fərdlərin davranışlarını yüksək dəqiqliklə, həm də onların razılığı olmadan izləməyə imkan verir. Bu imkanlar verilənlərin qorunmasının iki fundamental prinsipi (*verilənlərin gizliliyi və onlara əlyetərlik*) ilə ziddiyət təşkil edir. Elektron tibb məlumatlarından, yeni müayinə (*pasiyentin bədənindəki ötürücülərin köməyi ilə*) metodlarından istifadə tibb sahəsində irəli atılmış addımdır. Ancaq bu vəziyyətə əksər pasientlər son dərəcə həssaslıq göstərirlər. Belə ki, hətta mobil qurğular vasitəsilə anonim şəkildə həyata keçirilən böyük həcmli qeydiyyat verilənlərinin analizi zamanı bütün verilənlərin fərdi xarakteri itir, istifadəçinin (*pasiyentin*) şəxsi parametrlərini təyin etmək imkanı yaranır. İstifadəçinin şəbəkə parametrləri, məkan göstəriciləri və digər analiz üçün əlyetən olan məlumatlarının köməyi ilə çox böyük ehtimalla onun identikliyini müəyyən etmək mümkün olur.

Digər bir məsələ, kompüter texnologiyaları sahəsindəki təhlükəsizliklə bağlıdır (*kiber təhlükəsizlik*). İnformasiya hücumlarının təhlükə və riskləri "big data" fenomeni və informasiya sferasında mümkün cinayətlərin nəticələrinə adaptasiya olunmuş texniki həllər baxımından qiymətləndirilməlidir. İnformasiya təhlükəsizliyi sahəsindəki siyasət və prinsiplərə yenidən baxılmalıdır. Bu siyasət və prinsiplər də verilənlərin mühafizəsi qanununa ciddi riayət və şəxsi həyata hörmət etməyə yönəlməlidir.

Nəticə

Hazırda ekzabayt və zetabaytlarla BV axınının istehsalını təmin etmiş elm, texnika və texnologiyaların geniş yayıldığı erada yaşayırıq. Rəqəmsal informasiyanın həcmi artdıqca, bunlara əlyetərliyin və onlardan istifadə edən subyektlərin sayı da artır. "Big data" texnologiyalarının əsas vədi, tədqiqatçıların bacarıqlarından istifadə etməklə, həyatları xilas etmək, xidmətləri yaxşılaşdırmaq və dünyanı dərk etməkdən ibarətdir.

"Big data" analitika texnologiyalarını informasiya təhlükəsizliyi sistemlərində tətbiq etməklə milli təhlükəsizlik, terrorizm, qaçaqmalçılıq və s. kimi bir çox sahələrdə təhlükəsizlik risklərini azaltmaq mümkündür. Bu zaman istifadə edilən verilənlər veb, tibbi qurğular, sensorlar və s. vasitəsilə ötürülən məlumatlar ola bilər ki, bunların da bəziləri çox sadə (*hər kəs üçün açıq*) və bəziləri isə insanların şəxsi həyatını əks etdirən verilənlərdir. Bu verilənlər insanların sığorta ödənişlərinə, onlayn alqı-satqı ödənişlərinə təsir edə bilər. Verilənlərin əksəriyyəti fərdi məlumat olduğundan xüsusi mühafizə olunmalıdır. Fərdi məlumatların qorunması üzrə milli və beynəlxalq qanunvericiliyə uyğun xüsusi zəmanət olmalıdır. İnsanların razılığı olmadan onların şəxsi verilənlərinin analiz olunması yolverilməzdir. Ona görə də bu texnologiyaların köməyi ilə emal və analiz olunan informasiyanın təhlükəsizlik məsələləri isə istər tədqiqatçıların, istərsə də istehsalçıların diqqət mərkəzində olmalıdır.

"Big Data" texnologiyalarının müasir cəmiyyətə təsirini ümumiləşdirərək deyə bilərik ki, istənilən yeni texnologiya kimi, o da, "bütün dərdlərə dərman deyil"dir. Kifayət qədər güclü bir alət olsa da, müəyyən nöqsanları, məhdudiyyətləri və təhlükələri də vardır. Bütün bunlara baxmayaraq, "big data" bəşəriyyət üçün bir sıçrayış, elm və texnikadan başlayaraq biznesə qədər inkişafa doğru atılmış bir addımdır.

Ədəbiyyat

1. Əliquliyev R.M., Hacırəhimova M. Ş. "Big data" fenomeni: problemlər və imkanlar // İnformasiya Texnologiyaları Problemləri, 2014, №1, s.3–16.
2. The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Study report, IDC, December 2012. www.emc.com/leadership/digital-universe.
3. Diebold F.X. Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting. Discussion Read to the Eighth World Congress of the Econometric Society, 2000.
4. Dean J., Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters / Proceedings of the Sixth Symposium on Operating System Design and Implementation, vol.6, 2004, pp.137–150.
5. Hadoop. http://hadoop.apache.org/.
6. Alguliyev R., Imamverdiyev Y., Yusifov F. Some conceptual views on information security of the society // Journal of Communication and Computer, 2012, vol.9, pp. 644-648.
7. Lei X., Chunxiao J., Jian W., Jian Y., Yong R., Information Security in Big Data: Privacy and Data Mining // IEEE Access, 2014, vol.2, pp.1149–1176.
8. Big data and data protection.https://ico.org.uk/media/for-organisations/documents/1541/
9. Big data and privacy. A technological perspective. White House, May 2014.
10. Big data and privacy, MIT 2013. http://bigdata.csail.mit.edu
11. Alguliyev R., Imamverdiyev Y. Big Data: Big promises for information security / Proceedings of the IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), Astana, Kazakhstan 15-17 October, 2014, pp.216–219.
12. Laney D., 3D Data Management: Controlling Data Volume, Velocity and Variety. Technical report, META Group, Inc (now Gartner, Inc.), February 2001. http://blogs.gartner.com/
13. Gandomi A., Haider M. Beyond the hype: Big data concepts, methods, and analytics // Ted International Journal of Information Management, 2015, vol.35, pp.137–144.
14. Kambatla K., Kollias G., Kumar V., Grama A. Trends in Big Data Analytics // Parallel Distributed Computing, 2014, vol.74, no.7, pp.2561–2573.
15. Wu X., Zhu X., Wu G.Q., Ding W. Data mining with big data // IEEE Transactionson Knowledge and Data Engineering. 2014, vol.26, no.1, pp.97–107.
16. Che D., Safran M., Peng Z. From big data to big data mining: challenges, issues, and opportunities, in: B.Hong, X.Meng, L.Chen, W.Winiwarter, W.Song (Eds.), Database Systems for Advanced Applications, Springer, Berlin Heidelberg, 2013, pp.1–15.
17. http://cloudtimes.org/2014/02/12/gartner-report-big-data-will-revolutionize-the-cybersecurity-in-next-two-year/
18. Tene O., Polonetsky J. Big Data for All: Privacy and User Control in the age of analytics // Northwestern Journal of Technology and Intellectual Property, 2013, vol.11, no.5, pp.239–273.
19. Danger: 3 Reasons to Be Scared of Big Data. http://smartdatacollective.com/
20. Cloud Security Alliance, Expanded Top Ten Big Data Security and Privacy Challenges, 2013.
21. Malbrecht T., Prekopcsak Z. Big Data Security on Hadoop. www.rapidminer.com
22. "Fərdi məlumatlar haqqında" Azərbaycan Respublikasının Qanunu, 2010, 11 may. www.dmx.gov.az/userfiles/files/ferdimelumatlarqanun3.pdf
23. Duhigg C. How companies learn your secrets. New York Times, 2012, 16 February.
24. Agrawal D., Das S., Amr El Abbadi. Big Data and Cloud Computing: Current State and Future Opportunities / Proceedings of the 14th International Conference on Extending Database Technology, 2011, pp.530–533.
25. Yin C., Wang J., Xie C., et.al. Robot: An efficient model for big data storage systems based on erasure coding / IEEE International Conference on Big Data, 2013, pp.163–168.
26. Protection of personal data, 1995. http://eur-lex.europa.eu/legal-content/EN/

Гаджирагимова Макруфа Ш.
Институт Информационных Технологий НАНА, Баку, Азербайджан
makrufa@science.az

Технологии больших данных и проблемы информационной безопасности

Термин «Big Data», отражающий новую эру в обработке информации, широко обсуждается в средствах массовой информации, научной литературе и бизнес-среде. «Big Data» – это технология, способная внести революционные изменения в такие области, как управление, здравоохранение, наука, бизнес, коммерция, промышленность и др. С одной стороны, технологии открывают новые возможности для общества, но с другой стороны создают новые проблемы безопасности. В статье даётся краткое описание «Big Data» технологий. Исследуются польза аналитики «Big Data» и некоторые проблемы безопасности. Рассматриваются новые этические проблемы, созданные передовыми технологиями, с точки зрения защиты персональных данных, и предлагаются некоторые рекомендации.

Ключевые слова: *большие данные, анализ больших данных, безопасность, анонимный, персональные данные, идентификация, де-идентификация, шифрование.*

Makrufa S. Hajirahimova
Institute of Information Technology of ANAS, Baku, Azerbaijan
makrufa@science.az

Big data technologies and information security challenges

The notion of Big Data, which is a new era in information processing, is widely discussed in the business environment, mass media and scientific literature. Big Data technology can make revolutionary changes in the field of governance, health, science, business, e-commerce, industry and others. On the one hand, it opens up new opportunities for the society, but on the other hand, it causes new security problems. The article gives a brief description of the Big Data technology and investigates the use of Big Data analytics and some security issues. In addition, it considers the new ethical problems created by Big Data in terms of personal data protection and offers some recommendations.

Key words: *big data, big data analytics, security, anonymisation, privacy, identification, de-identification, encryption.*