“Big Data” Təhlükəsizliyi və Həlli yolları

Tural Yunusov
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
turaly@mail.ru

***Xülasə* — Məqalə “Big Data” təhlükəsizliyi və həlli yollarına həsr edilmişdir. Məqalədə “Big Data” fenomeni, “Big Data” texnologiyası, verilənlərin təhlükəsizlik məsələləri, o cümlədən verilənlərin idarəedilməsi, icazələrin idarəedilməsi, məlumatın mühafizəsi və konfidensiallığı, şəbəkə təhlükəsizliyi analiz edilmiş və mövcud problemlər müəyyən edilmişdir.**

Açar sözləri— “Big Data”; “Big Data” Təhlükəsizliyi; “Big Data” təhlükəsizlik meyarları; “Big Data”-nın emal edilməsi; konfidensiallıq; DLP sistemlər

I. GİRİŞ

Keçmişdən günümüzə qədər gələn müddət ərzində məlumat gündən-günə artmış, hətta son illərdə İnformasiya Texnologiyalarının sürətli inkişafı ilə əlaqədar dəfələrlə çoxalmışdır. Bunun nəticəsində "informasiya çirkliliyi" adlanan termin yaranmışdır. Bir çox şirkətlər bu mövzu ilə əlaqədar araşdırma apardıqdan sonra “Big Data” (Böyük Verilənlər) termini yaranmışdır.

“Big Data” dedikdə müxtəlif yollarla əldə edilən informasiyaların (şəbəkə, bloq, şəkil, video, log faylları və s.) yığılaraq yaratdıqları böyük həcmdə və normal alətlərlə emal oluna bilməyən verilənlər toplusu nəzərdə tutulur. Bu yeni termin həcm və mürəkkəblik baxımından mövcud idarəetmə metodları və intellektual analiz vasitələri ilə emal oluna bilməyən verilənləri təyin etmək üçün istifadə edilir.

İnkişaf etməkdə olan İnformasiya Cəmiyyəti böyük elmi-texniki inqilab çağındadır və bu inqilab biznesdən başlayaraq elmə qədər həyatımızın bütün sahələrinə nüfuz etməkdədir. Bu eyni zamanda sosial-iqtisadi, təhlükəsizlik elmi baxımdan mürəkkəb bir problemdir. Problem böyük verilənlər ideyası ilə daha da dərinləşməkdədir. Belə ki, “Big Data” saxlanması, idarə edilməsi və onlardan informasiya əldə edilməsi ciddi problem yaratmışdır. Problem müxtəlif mənbələrdən avtomatik və fasiləsiz olaraq generasiya olunan verilənlərin real-vaxt ərzində emalı və analizində mövcud İT (İnformasiya Texnologiyaları) həllərin səmərəsiz olmasındadır. Mövcud vəziyyət “verilənləri yaratmaq son dərəcə asan, emal etmək isə son dərəcə çətin olur” fikrini söyləməyə əsas verir [1].

“Big Data” texnologiyalarının təqdim etdiyi imkanlardan informasiya təhlükəsizliyi problemlərinin həlli üçün geniş istifadə etmək mümkündür. İnformasiya təhlükəsizliyi və fərdi məlumatların qorunması “Big Data” üçün həlli vacib məsələlərdəndir. Bu problemlərə böyük həcmdə verilənlərin idarə edilməsi infrastrukturunda informasiya təhlükəsizliyi, verilənlərin konfidensiallığı və fərdi məlumatların qorunması, girişə nəzarət, audit və informasiya təhlükəsizliyinin real zaman rejimində monitorinqi və s. daxildir [2].

“Big Data” texnologiyaları informasiya təhlükəsizliyi insidentlərinin avtomatik cavablandırılması, onların nəticələrinin qısa müddətdə, minimal xərclərlə aradan qaldırılması və daha keyfiyyətli tədqiqatın aparılmasında xüsusi rol oynayır [3].

II. BİG DATA TEXNOLOGİYALARI

“Big Data” platformasının meydana gəlməsində beş komponent iştirak edir. Bunlar həcm (volume), sürət (velocity), müxtəliflik (variety), həqiqilik (veracity) və dəyər (value) –dən ibarətdir. İngilis dilli mənbələrdə bunu «5V» də adlandırırlar [4].

Müxtəliflik (variety): Emal olunan informasiyanın 80% -i strukturlaşmış (strukturlaşmış - verilənlər bazaları ilə əlaqəli olan) informasiya deyildir və hər yeni texnologiya müxtəlif tiplərdə informasiya emal edir. Telefonlardan, planşetlərdən, birləşmiş şəbəkələrdən müxtəlif tip informasiya emal edilir və bu informasiyaların fərqli dillərdə ola biləcəyi nəzərə alınmalıdır, həmçinin bu verilənlərin bir-birlərinin tipinə çevrilmələri də vacibdir.

Sürət (Velocity): Böyük verilənlərin emal sürəti çox yüksəkdir və ildən ilə bu sürət artmaqda davam edir. Sürətlə artan verilənlər, əməliyyatların sayının və müxtəlifliyinin də eyni sürətlə artmasına gətirib çıxarır.

Həcm (Volume): IDC (International Data Corporation) hesabatına əsasən 2020-ci ildə informasiyanın həcmi 2009-cu ildəkindən 44 dəfə çox olacaq. Hal-hazırda istifadə edilən, "böyük" adlandırdığımız məlumat saxlama yaddaşları və "böyük sistemlər" gələcəkdə sürətlə artan verilənlər üçün kifayət etməyəcək [5, 6].

Həqiqilik (Veracity): İnformasiya axını zamanı verilənlərin “etibarlı” şəkildə təşkil olunması vacib məsələdir. Verilənlərin təhlükəsizlik səviyyəsinin monitorinqi, səlahiyyətli şəxslər tərəfindən istifadəsi və ya gizli qalması öz vacibliyini qoruyur.

Dəyər (Value): “Big Data” platformasının ən əhəmiyyətli komponenti isə dəyər yaratmasıdır. Bütün yuxarıda sadalananlar “Big Data” komponentlərinin emalından sonra təşkilat üçün bir müsbət dəyər yaratmalıdır. Məsələn:

Müəyyən qərarlar verərək yoluxucu xəstəliklərin qarşısını almaq üçün səhiyyə mövzusunda strateji qərarlar verən bir dövlət təşkilatının bölgə, rayon, xəstəlik, dərman, həkim məlumatlarını əldə etməlidir. Digər bir misal, istənilən bank kredit verəcəyi adamın yalnız demoqrafik məlumatlarını deyil, onun tətil etmə vərdişlərini belə izləyə bilməli, lazım gələrsə sosial şəbəkələrdə nə etdiyini görə bilməlidir.

“Big Data”-nın emal edilməsi

Hal-hazırda istifadə olunan verilənlər bazaları inkişaf etsə də, böyük verilənləri emal etmək üçün bizə lazım olan imkanları vermir. Çünki, əldə edilən məlumatlar bir-birləri ilə əlaqəli olmur, bu səbəbdən də bazalarla əlaqə yaratmaq üçün tələb olunan SQL sorğularını yazmaq və böyük məlumatları analiz etmək çətinlik yaradır. Bununla birlikdə əldə etdiyimiz məlumatlar strukturlaşmış olmur, yəni strukturlaşdırılması üçün yenidən baxılır. Bizə lazım olan məlumatları əldə etməyimiz üçün klassik üsullardan daha fərqli həllər lazımdır [7]. Məsələn, Google özü yaratdığı texnologiya əsasında məlumatları yığır (Google File System), emal edir (MapReduce) və saxlayır (Big Table). Eyni şəkildə Amazon da özü yaratdığı texnologiya (DynamoDB) əsasında məlumatları yığır. Bu şirkətlər yaratdıqları texnologiyalar haqqında yalnız elmi yazılar dərc etdirir və bir çox proqramçılar bu texnologiyaları inkişaf etdirir, məsələn, Hadoop, HBase, Lucene kimi texnologiyaları [8]. Facebook, Twitter, LinkedIn kimi şirkətlər isə daha inkişaf etmiş texnologiyalar təklif edir, məsələn,. Cassandra, Hive, Pig, Voldemort, Storm kimi texnologiyalar.

III. BİG DATA TƏHLÜKƏSİZLİYİ MEYARLARI

Big Data Təhlükəsizliyin meyarları aşağıdakıladan ibarətdir [9].

1. Verilənlərin idarə edilməsi
2. İcazələrin idarə edilməsi
3. Məlumatın mühafizəsi və konfidensiallıq
4. Şəbəkə təhlükəsizliyi

“Big Data” təhlükəsizliyi kriterilərinə yuxarıda sadalanan 5 əsas təhlükəsizlik riski və təhlükəsizlik riskinin azaldılması üçün zəruri olan 21 alt komponent daxildir. Ümumi təhlükəsizlik kriteriləri aşağıdakılardır [10]:

– *Verilənlərin idarə edilməsi.* Verilənlərin idarə edilməsi üç əsas qrupa bölünür: verilənlərin təsnifatı, verilənlərin aşkarlanması və verilənlərin nişanlanması.

Verilənlərin təsnifatı “Big Data” platformasında effektiv təhlükəsizlik nəzarətinin həyata keçirilməsində əhəmiyyətli rol oynayır. Çox vaxt təşkilatlarda böyük həcmdə verilənlər yığılır. Verilənlərin hansı mövzularda olduğunu aydınlaşdırmaq, onlar arasında hansı səviyyədə kriptoqrafiq mühafizənin təmin edilməsi lazım gəlir.

Müəyyən situasiyalarda həsas verilənlərin aşkarlanması təşkilatı bir çox risklərdən azad edir. Verilənlərin aşkarlanması iki formada yerinə yetirilir: strukturlaşmış verilənlər üçün və strukturlaşmamış verilənlər üçün.

Strukturlaşmış verilənlər – verilənlər bazaları ilə əlaqəli olan verilənlərdir. Buna misal olaraq Hadoop (Böyük verilənlərin de-fakto standartı hesab olunan Apache Software Fondation-un layihəsi *Hadoop* paylanmış hesablama mühitində böyük verilənlərin emalı və analizi üçün əsas platformadır) üçün verilənlərin yerini və təsnifatını təyin edən, verilənlər bazası ilə məlumatların əlaqələndirilməsi kimi *comma-separated values* (CSV) və ya *JavaScript Object Notation* (JSON) fayl tiplərini göstərmək olar. Bu səbəbdən sütunların və xanaların təhlükəsizliyi proqramlar vasitəsi ilə həll edilir. Məsələn mütəxəssis tərəfindən sütun və sətirlər müəyyən təhlükəsizlik səviyyəsində gizlədilə və ya göstərilə bilər [11].

Strukturlaşmamış verilənlər – bir-biri ilə və verilənlər bazaları ilə əlaqəli olmayan verilənlərdir**.** Strukturlaşmamış verilənlər üçün təsnifatın aparılması və emal edilərək həsas verilənlərin çıxardılması daha çətindir [12].

– *İcazələrin idarə edilməsi*. Verilənlər bazaları və ona olan icazələr mütləq şəkildə idarə olunmalıdır. Əgər istifadəçilər daha çox səlahiyyətlərə malik olarsa, bu verilənlərin dəyişdirilməsi və ya silinməsinə gətirib çıxarar. İcazələrin idarə edilməsi üçün istifadəçi avtorizasiyası qaydaları yaradılmalı, mərkəzləşdirilmiş icazələrin idarə edilməsi siyasəti ilə istifadəçilərə məlumatlara icazələr təyin edilməlidir.

– *Məlumatın mühafizəsi və konfidensiallığı*. Verilənlərin mühafizəsi və konfidensiallığı böyük verilənlərin təhlükəsizliyində çox böyük əhəmiyyətli rol oynayır. Məlumatın mühafizəsi və konfidensiallığı iki səviyyədə həyata keçirilir: Tətbiq səviyyəsində şifrlənmə – verilənlər proqram təminatı səviyyəsində şifrlənir; disk səviyyəsində şifrləmə – burada verilənlərin yerləşdiyi yaddaş qurğuları bütövlüklə şifrlənir.

– *Şəbəkə təhlükəsizliyi.* Böyük verilənlərin yığılması, analizi və emalında şəbəkədən istifadə olunur. Bu səbəbdən verilənlərin şəbəkə vasitəsi ilə ötürülməsi zamanı təhlükəsizlik təmin edilməlidir.

IV. VERİİLƏNLƏRİN TƏHLÜKƏSİZLİK HƏLLƏRİ

Böyük verilənlərin təhlükəsizlik problemlərinin həlli üçün şirkətlər müxtəlif praktiki üsullardan istifadə edir. Bu həllərdən biri də (DLP – Data Loss Prevention) verilənlərin itkisinin qarşının alınması sistemlərinin istifadəsidir [13]. DLP sistemlər həssas verilənlərin şəbəkə daxilində icazəsiz istifadəsini və monitorinqini apararaq təhlükəsizliyini təmin edir. DLP sisteminin şəbəkəyə qoşulma sxemi aşağıdakı kimidir (şəkil.1).

Şəkil 1. DLP sisteminin şəbəkəyə qoşulma sxemi

DLP sistemlərində verilənlərin tipləri

İnformasiya təhlükəsizliyində “verilənlərin təhlükəsizliyi” kateqoriyasında qiymətləndirilən verilənlərin itkisinin qarşısının alınması texnologiyasının məqsədi verilənlərin arxivdən istifadəçiyə təhlükəsiz şəkildə çatdırılmasını təmin etməkdir. Bu məqsədlə DLP sistemlər üç səviyyədə təhlükəsizliyi təmin edir [14]:

– *Hərəkətli verilənlər (Data in Motion).* Şəbəkə daxilində hərəkət edən, yəni e-poçt, ismariclər, veb və P2P (Peer-to-peer) ötürmə kanalları üzərindən daimi hərəkət edən verilənlər tipidir.

– *İstifadə olunmayan verilənlər (Data at Rest).* Verilənlər bazası, qovluq sistemləri və digər xüsusi arxiv bölmələrində saxlanılan, əsasən ilk addımda təhlükəsizliyinin təmin olunması vacib olmayan verilənlər tipidir.

– *İstifadə olunan verilənlər (Data in Use):* Son istifadəçinin daimi olaraq istifadə etdiyi və verilənlər bazası ilə əlaqəli olan aktiv verilənlər tipidir.

NƏTİCƏ

“Big Data” böyük informasiya massivlərindən istifadə etməyə imkan verən yeni nəsil texnologiyadır. Bu texnologiyadan düzgün istifadə müxtəlif fəaliyyət sahələrində iş prossesinin sürətlənməsinə gətirib çıxarır. Böyük həcm, sürət və müxtəliflik kimi xüsusiyyətlərlə xarakterizə olunan verilənlərin emalı və analizi məqsədilə MapReduce, Hadoop, HDFS, NoSQL və s. proqram-aparat platformaları yaradılmışdır.

“Big Data” texnologiyalarının təhlükəsizlik sahəsində istifadəsi təhdidlərin və risklərin erkən aşkar edilməsinə inkan verər. Bunla yanaşı “Big Data”-nın emal prosesi zamanı təhlükəsizliyinin təmin edilməsi çox vacib məsələdir və məsələnin həlli üçün yeni metodların işlənməsi aktualdır. Böyük həcmli verilənlərin emalı sahəsində superkompüterlərin, qrid və bulud texnologiyaların istifadəsi ilə müəyyən problemlər aradan qaldırılsa da, bu sahədə hələ də ciddi texniki və elmi-nəzəri problemlər möcuddur. Problem ancaq böyük həcmdə verilənlərin saxlanılması və idarə olunmasında deyil, həm də strukturlaşdırılmamış verilənlərin analizi və nəticələrin interpretasiyasındadır.

ƏDƏBİYYAT

[1] R. Əliquliyev, M. Hacırəhmanova, “Big Data” Fenomeni: problemlər və imkanlar // İnformasiya texnologiyaları problemləri jurnalı, 2014, №2, 3-16

[2] Big Data, Big Impact: New Possibilities for International Development, world economic forum, 2012, *www.weforum.org*

[3] EMC Big Data 2020 Projects, http://www.emc.com/leadership/digital-universe/iview/big-data-2020.htm

[4] Oracle and Big Data: Big Data for the Enterprise, 2013, http://www.oracle.com/

[5] Big Data Solutions, 2013, http://www8.hp.com/

[6] Big Data, 2013, http://www.microsoft.com/.

[7] D. Agrawal, S. Das, A. Abbadi, Big Data and Cloud Computing: Current State and Future Opportunities, EDBT, Uppsala, Sweden, march 22–24, 2011.

[8] Hadoop Distributed File System, http://hadoop.apache.org/docs/

[9] Securing Big Data Security issues with Hadoop environments, Securosis, https://securosis.com/blog/securing-big-data-security-issues-with-hadoop-environments

[10] Ways to Build a Better Big Data Security Strategy, http://www8.hp.com/h20195/V2/GetPDF.aspx/4AA5-0863ENW.pdf

[11] Big Data Security Assessment, http://www.intel.ru/content/dam/www/public/us/en/documents/datasheets/big-data-security-assessment-datasheet.pdf

[12] Big Data Analytics for Security Intelligence, https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Analytics_for_Security_Intelligence.pdf

[13] T. Smith, Big Data Security: The Evolution of Hadoop’s Security Model, http://www.infoq.com/articles/HadoopSecurityModel

[14] Data loss prevention - Ernst & Young, http://www.ey.com/Publication/vwLUAssets/EY_Data_Loss_Prevention/$FILE/EY_Data_Loss_Prevention.pdf

[15] Symantec Data Loss Prevention Solution, https://www.symantec.com/content/en/us/enterprise/fact_sheets/data-loss-prevention-solution-ds-21350666.pdf