Zaman sıralarında anomaliyaların aşkarlanması metodlarının analizi

Afət Mikayılova
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
afet_mikayilova@mail.ru

***Xülasə*— Bu məqalədə şəbəkə trafikinin müasir dövrdə aktuallığı analiz edilmişdir. Şəbəkə trafikinin yaratdığı və son dövrlərdə rast gəlinən problemləri ilə bağlı tədqiqat aparılmışdır. Zaman sıralarında mövcud olan anomaliyaların aşkarlanması üçün təklif edilmiş mövcud metodların analizi aparılmışdır. Metodların məqsədi və problemləri haqqında məlumatlar qeyd edilmişdir.**

Açar sözlər— Şəbəkə trafikinin analizi, zaman sıraları, anomaliya, oxşarlıq əsaslı metod, pəncərə əsaslı metod, proqnozlaşdırma əsaslı metod, gizli markov modeli

I. ŞƏBƏKƏ TRAFİKİNİN ANALİZİ

Şəbəkə trafiki monitorinqi üçün şəbəkə trafiki analizi hal-hazırda daha da vacib məsələyə çevrilmişdir. Ötən illərdə administratorlar yalnız az sayda şəbəkə qurğusunun monitorinqini aparırdılar. Şəbəkənin genişliyi yalnız az və ya 100 Mbts ola bilər. Hal-hazırda administratorlar daha yüksək sürətli kabel şəbəkəsi (saniyədə 1 Gbit / s-dən çox) və Asinxron Transfer Rejimi (Asinxron Transfer Mode, ATM) şəbəkələri və simsiz şəbəkələr kimi müxtəlif şəbəkələrlə məşğul olmalıdırlar. Şəbəkəni idarə etmək, çatışmazlığını aradan qaldırmaq, təhlükəsizliyini təmin etmək və problemlərini surətli həll etmək üçün əlavə şəbəkə trafikinin analizi tələb olunur. Şəbəkə trafikinin analizi son dövrlərdə bir sıra çətinliklər yaradır. Şəbəkə müxtəlif səviyyələrdə təhlil edilir. Bu səviyyələr təhlükəsizlik, paket, kanal və şəbəkə səviyyəsinə aid edilir. Trafik təhlili üçün tədqiqatçılar müxtəlif üsullardan istifadə edirlər. Şəbəkə trafikinin analizi əvvəlcədən işləmə və şəbəkə məlumatlarından nümunələri aşkar etmək üçün faktiki analiz və müşahidələrdən istifadə edilir. Şəbəkə trafiki analizinin üç əsas mərhələsi ilkin proseslər üsulu, analiz (bu mərhələdə Data mining (DM) üsulundan istifadə edilir) və qiymətləndirmədən ibarətdir [4]. İlkin prosesler üsulu şəbəkə trafikini manipulyasiya etmək üçün mühüm mərhələdir və real məlumatları başa düşülən bir formaya çevirir. DM biliklərin aşkarlanması üçün istifadə olunur və şəbəkə trafiki analizində mühüm rol oynayır.

Şəbəkə trafikinin proqnozlaşdırılması əhəmiyyətli bir məsələdir. Şəbəkə trafikini proqnozlaşdırılması–şəbəkənin monitorinqi, şəbəkə təhlükəsizliyi, şəbəkə trafikinin qarşısının alınması və şəbəkələrin sürətini artırılması üçün mühüm məsələlərdən birdir [7].

II. MÖVCUD METODLARIN ANALİZİ

Zaman sıralarında anomal halların aşkarlanması üçün tədqiqatçılar bir çox metodlar təklif etmişdir. Aşağıdakı üsullar tədqiq edilmişdir:

- Oxşarlıq əsaslı metod
- Pəncərə əsaslı metod
- Praqnozlaşdırma əsaslı metod
- Gizli Markov Modeli

Oxşarlıq əsaslı metodla test zaman seriyasının anomaliyaların dərəcisini hesablamaq üçün müvafiq məsafə və ya oxşarlıq əlaqəsindən istifadə edərək, test və təlim zaman sıralarının arasında bu uyğunluqdan istifadə olunur [2]. Zaman sırasındakı anomaliyalar digər anomaliyalardan fərqli olur və bu fərq yaxınlıq ölçüsü (proximity measure) kimi istifadə edilə bilir. Beləliklə, Oxşarlıq əsaslı metod zaman sıralarının hər hissəsi üçün yaxınlığın ölçüsünü müəyyən etmək üçün istifadə edilir [5]. Bu üsulun əsas məqsədi yaxın qonşuları qiymətləndirməkdir, davamlı vaxt seriyası üçün ən yaxın qonşu (K-nearest neighbor for continuous, KNNC) üsulu adlanır və yaxınlığın ölçüsünü ifadə edir. KNNC test zaman sırasında anomaliya hesabını təyin edir və təlim verilənlər bazasında ən yaxın qonşusuna olan məsafəyə bərabər olur. Metodun əsas çatışmazlığı odur ki, bütün zaman seriyasının anomal olub olmadığını müəyyən edə bilir, lakin anomal sonluğu dəqiq tapa bilmir. Anomaliyaya səbəb olan zaman seriyasında dəqiq bölgəni müəyyən etmək üçün zaman seriyasını emal etmək lazımdır. Beləliklə, bu üsulun icrası çox vaxt seçilməsi asan olmayan yaxınlıq ölçüsündən çox asılıdır.

Pəncərə əsaslı metod ilə anomaliyanı aşkar etmək üçün verilən zaman seriyasını sabit ölçülü pəncərələrə bölür, bir və ya daha çox pəncərədə anomaliya səbəbini müəyyən etmək üçün istifadə olunur. Bütün zaman sıraları kiçik nəticələrə bölündükdən sonra, biz anomaliyanın olub-olmadığını asanlıqla müəyyən edə bilərik. Bu üsul test zaman sırası aralığından sabit uzunluqlu pəncərələri çıxarır və hər bir pəncərəyə anomaliya ölçüsü təyin edir [1]. Sonra ilkin pəncərə hesabları, test zaman seriyası üçün anomaliya hesabını əldə etmək üçün toplanır. Bir pəncərə hesabının təyin edilməsi və hesabların birləşməsi prosesləri müxtəlif yollarla edilə bilər [3]. Pəncərə əsaslı üsulun əsas çatışmazlığı pəncərənin ölçüsünün diqqətlə seçilməsidir ki, o, anomaliyi açıq şəkildə

DOI: 10.25045/NCInfoSec.2018.46

əldə edə bilsin. Pəncənin optimal ölçüsü anormal zaman sırasındakı anormal bölgənin uzunluğundan asılıdır. Pəncərə əsaslı texnikanın digər çatışmazlığı hesablamalara görə bahalı olmasıdır.

Proqnozlaşdırma əsaslı metod proqnozlaşdırıcı modellərdən istifadə edərək anomaliyanın aşkarlanması üçün araşdırılmışdır. Bu üsul, statistik proseslərdə normal zaman sıralarını müəyyənləşdirmək üçün yaradılımışdır və əgər prosesdə hər hansı anomaliya baş verərsə bu metod işləməyəcəkdir [2]. Beləliklə, əsas məsələ bundan ibarətdir ki, normal zaman sıralarının təlim verilənlər bazasındakı proseslərin parametrlərini öyrənmək və daha sonra öyrənilmiş prosesdən test zaman sıralarının yaranma ehtimalını qiymətləndirməkdir. Bütün proqnozlaşdırmaya əsaslanan metodlar sabit uzunluqlu məlumatlardan istifadə edir. Eyni zaman sırasında, bəzən daha kiçik bir məlumatı proqnozlaşdırır, lakin digər vaxtlarda daha uzun bir məlumata ehtiyacı olur. Beləliklə, dinamik uzunluqlu məlumatdan o halda istifadə etmək olar ki, əgər bu üsul hər hansı bir müşahidədə m uzunluğunda hesabatların verildiyi prosesləri yüksək təhlükəsizliyə görə proqnozlaşdırıla bilməyəcəyi təqdirdə, müşahidənin daha yüksək etibarlılığını proqnozlaşdırmaq üçün hesabatın uzunluğunu artıra və ya azalda bilsin. Proqnozlaşdırma əsaslı üsullar hər bir müşahidə üçün anomaliya ölçüsünü hesablayır [4]. Beləliklə, onlar hər cür anomaliyaları əhatə edə bilərlər: zaman seriyasındakı anomal müşahidələri, zaman seriyasındakı anomal ardıcıllıqları, zaman sırasındakı ümumi anomaliyaları və s.

Gizli Markov əsaslı model (HMM) müşahidə edilə bilən parametrlərdən istifadə edərək bir sistemin xarakterizə etdiyi güclü sonlu üsuldur [6]. Bu üsul ardıcıl modelləşdirmə və ardıcıl anomaliyanın aşkarlanması üçün geniş istifadə olunur. Həmçinin zaman seriyasındakı anomaliyanın aşkarlanmasına tətbiq edilir [8]. Gizli Markov əsaslı model proseslərdə, müəyyən bir zaman seriyası $O=O_1 \ldots O_n$ kimi göstərilə bilər. $Q= Q_1 \ldots\ldots Q_n$ bir əsas (gizli) zaman seriyaları isə dolayı yolla müşahidə olunur. Əslində, real zaman seriyasını meydana gətirən proseslər bu cür müşahidə olunmaya bilər, baxmayaraq ki, gizli zaman sıralarını yaradan proses gizli Markov əsaslı üsuldur. Beləliklə, normal zaman sıraları HMM istifadə edərək modelləşdirilə bilər, anomal zaman sıraları isə qeyri-mümkündür. HMM əsaslı üsulun əsas çatışmazlığı ondan ibarətdir ki, gizli Markov əsaslı model mövcud gizlin proseslerdə baş verir və normal zaman sıraları yaradır. Gizli Markov əsaslı üsulun olmadığı təqdirdə, bu üsullar anomaliyaları aşkar edə bilmir. HMM üsulu, zaman sırası üçün markov modelini qurur [8]. Beləliklə, hər növ anomaliyaların aşkarlanmasına kömək edən zaman sıralarındakı hər bir müşahidə üçün ehtimal olunan müddətdə anomal prosesləri qiymətləndirir.

Beləliklə, biz yuxarda qeyd etdiyimiz metodların məqsədini ümumiləşdirərək aşağıdakı cədvəldə göstərə bilərik:

CƏDVƏL 1.

Sıra	İstinad	Metod	Məqsəd
1.	[2]	Oxşarlıq əsaslı metod	Zaman sıralarının hər hissəsi üçün yaxınlığın ölçüsünü müəyyən etmək
2.	[3]	Pəncərə əsaslı metod	Bütün zaman sıraları kiçik pəncərələrə bölündükdən sonra, anomaliyanın olub-olmadığını asanlıqla müəyyən etmək
3.	[4]	Praqnozlaşdırma əsaslı metod	Proqnozlaşdırıcı modellərdən istifadə edərək anomaliyanı aşkarlanmaq
4.	[8]	Gizli Markov metodu	Müşahidə edilə bilən parametrlərdən istifadə edərək zaman seriyasındakı anomaliyanı aşkarlanmaq

ƏDƏBİYYAT

[1] V. Chandola, A. Banerjee, and V. Kumar. “Anomaly detection: A survey,” ACM Computing Surveys (CSUR), 2009, vol. 41, no. 3, pp. 1-72.

[2] B. Pincombe, “Anomaly detection in time series of graphs using ARMA processes,” Asor Bulletin, 2005, vol. 24, no. 4, pp. 2-10.

[3] V. Chandola, D. Cheboli, and V. Kumar, “Detecting anomalies in a time series database,” Technical Report 09-004, 2009, pp. 1-12.

[4] M. Joshi, T.H. Hadi, “A Review of Network Traffic Analysis and Prediction Techniques, 2015,” arXiv:1507.05722 [cs.NI], p. 23.

[5] T. Lotze, G. Shmueli, S. Murphy, and H. Burkom, “A wavelet-based anomaly detector for early detection of disease outbreaks,” Proceedings of the 23nd International Conference on Machine Learning, 2006, pp. 1-6.

[6] V. Jecheva, “About some applications of hidden markov model in intrusion detection systems,” In International Conference on Computer Systems and Technologies (CompSysTech), 2006, pp. 1-6.

[7] Y. Yu, M. Song, Z. Ren, l. Song, “Network Traffic Analysis and Prediction Based on APM,” Proc. of the IEEE 6th International Conference on Pervasive Computing and Applications, pp. 275-280.

[8] Z. Liu, J.X. Yu, L. Chen, and D. Wu, “Detection of shape anomalies: A probabilistic approach using hidden markov models,” Proc. of the IEEE 24th International Conference on Data Engineering, 2008, pp. 1325-1327.

ANALYSIS OF METHODS FOR ANOMALY DETECTION IN TIME SERIES

Afet Mikayilova

Institute of Information Technology , Baku, Azerbaijan

afet_mikayilova@mail.ru

Abstract –This article was analyzed the relevance of network traffic in modern times. It has been investigated about network traffic caused shortcomings which observed in the recent years. It has been analysed available methods for the detection of existing anomalies in the time series was performed. it was noted information about objectives and disadvantages of methods.

Keywords – network traffic analysis, time series, anomaly, proximity based method, window based method, prediction based method, hidden markov method.