DOI: https://doi.org/10.26565/2304-6201-2024-64-04

УДК (UDC) 519.216

Danilevskyi Mykhailo *PhD student; V.N. Karazin Kharkiv National University, Svobody Square, 4, Kharkiv-22, Ukraine, 61022.*
e-mail: m.danilevskiy@gmail.com
https://orcid.org/0009-0000-0030-2218

Yanovsky Volodymyr *Doctor of Physical and Mathematical Sciences, professor; V. N. Karazin Kharkiv National University, sq. Svobody 4, Kharkiv, Ukraine, 61000; Institute of Single Crystals, National Academy of Sciences of Ukraine, Nauki Ave. 60, Kharkiv, Ukraine, 61001*
e-mail: yanovsky@isc.kharkov.ua
https://orcid.org/0000-0003-0461-749X

Detecting Telephone Subscribers with Abnormal Behaviour Through Network Properties Analysis

Abstract. The use of telephone subscriber networks for fraud, sales of goods and services, and spam leads to annual financial losses of billions of dollars worldwide. The problem was studied back in 1996 and is still relevant today, despite many methods of counteraction and protection. Traditional methods, such as blocking lists and call frequency monitoring, are often ineffective against spammers who bypass these systems by changing their behaviour patterns. **Objective.** The purpose of the work is to study the use of subscriber network properties, such as clustering coefficients, centrality, and average shortest path length, as criteria for identifying subscribers with abnormal behaviour in a dynamic telephone network. **Research methods.** Modeling and numerical experiment. **Results.** The study shows that the global clustering coefficient is a sensitive measure for detecting the presence of spammers in the network. Its value decreases significantly when spammers appear. When classifying subscribers into normal and spammer using the Random Forest model, the most important properties are the local clustering coefficient, average shortest path length, degree and centrality of the subscriber in the network. According to the telephone network modeling data, it was found that with a measurement window size of 4 days, the classifier accuracy (F1 score) and the accuracy of detecting spammers (TPR) reached values of 80%. **Conclusions.** Using the network characteristics of subscribers has a positive effect on the accuracy of detecting subscribers with abnormal behaviour, but it takes time for spammers and normal subscribers to become distinguishable by network characteristics.

Keywords: *malicious phone calls, mobile call graph, telephone spam detection, telephone network, lognormal distribution, degree distribution, clustering coefficient, average shortest path length*

Як цитувати: Danilevskyi M., Yanovsky V. Detecting Telephone Subscribers with Abnormal Behaviour Through Network Properties Analysis. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2024. вип. 64. С.32-39. https://doi.org/10.26565/2304-6201-2024-64-04

How to quote: M. Danilevskyi, V. Yanovsky "Detecting Telephone Subscribers with Abnormal Behaviour Through Network Properties Analysis", *Bulletin of V. N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 64, pp. 32-39, 2024. https://doi.org/10.26565/2304-6201-2024-64-04

1 Introduction

Malicious phone calls, encompassing both spam and scams, pose a significant global problem, resulting in billions of dollars in financial losses each year. This issue has been a subject of study since at least 1996 [1] and remains pressing today. Mobile phone spam involves the exploitation of telecommunications networks to execute fraudulent or distracting activities. It is particularly problematic for network providers and users, leading to substantial financial losses and damages. One of the critical challenges in addressing mobile phone spam is the dynamic nature of malicious caller behaviour, which frequently changes and adapts to detection efforts. Although various countermeasures have been implemented, a comprehensive and effective solution to eradicate these activities has yet to be developed. For example, countries such as the United States have introduced measures like the National Do Not Call Registry [2] to mitigate the issue, with fines for violations reaching millions of dollars [3].

© Danilevskyi M., Yanovsky V., 2024
This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0.

Numerous technical approaches have been developed to identify users with abnormal activity within social and, specifically, telephone networks who engage in fraudulent activities or disturb normal users through unsolicited advertisements [4-14]. One common method involves gathering reports from users via mobile applications to generate "blacklists" of known offenders [15]. Another approach includes tracking user behavioural characteristics, such as the number of calls, call timing, duration, and the occurrence of international calls, to create user profiles based on these data points [1, 2, 16]. Additionally, methods of call content analysis – examining both spoken words [17] and the acoustic properties of the voice [18] – have been employed to detect malicious subscribers. Topological features, such as the degree of connectivity, the number of groups a user participates in, interconnections among friends, and the average number of connections within a user's community, have also been utilized in identifying suspicious network users, as demonstrated by the authors of [19]. Another technique proposed in [20] uses the Graph-Based Anomaly Detection (GBAD) system to represent call data as a graph, thus enabling the identification of anomalous structures that deviate from the patterns of normal users. In [21] and [22], neural networks were applied to detect users exhibiting anomalous behaviour. Despite the diverse strategies for identifying malicious subscribers, the issue remains significant and unresolved.

In this paper, we examined potential criteria for identifying subscribers with abnormal behaviour within a mobile telephone network. A computer simulation was developed to replicate a dynamic network of telephone subscribers. The modeling data were used to analyze various properties that characterize subscriber behaviour within the network. Among the indicators evaluated were degree distribution, clustering coefficient, betweenness centrality, average shortest path length, and others. These modeling results facilitated the identification and assessment of network properties that define phone user behaviour. The findings indicate that the global clustering coefficient is effective in detecting the presence of spammers, while properties such as local clustering coefficient, average shortest path length, degree, and betweenness centrality are crucial for subscriber classification.

This approach enables malicious users who may evade traditional spam filters based on call frequency or block lists, including those who have remained inactive for extended periods, frequently change numbers, or exhibit irregular behaviour patterns.

2 Modeling

In this study, we employ a modified version of the dynamic network model of telephone subscribers, as compared to the model described in [27]. The network model consists of nodes representing subscribers and edges representing calls or links between them. A link is established whenever two subscribers engage in a call. The model accounts for variations in call frequency, with typical subscribers making 5 to 7 calls per day, while users exhibiting higher calling activity, such as spammers, may place between 200 and 300 calls daily. The network evolves dynamically over time, with new connections forming and existing connections breaking based on ongoing and completed calls. The dataset generated from the model provides information about calls over a period, enabling an analysis of user properties. In this model, subscribers initiate calls to others on their contact list with a probability defined by a model parameter. Normal users' contact lists contain other normal telephone subscribers such as family members, friends, colleagues etc. Spammers do not have such contact lists and make random calls to a wide range of users. The number of calls follows a lognormal distribution [23]. Additionally, the model assumes that subscribers cannot make or receive other calls during an ongoing call. The duration of each call is determined by the probability of its termination at any given moment, which is also defined by a model parameter.

The model parameters include the total number of subscribers, the parameters of the lognormal distribution for simulating the daily number of calls, the duration of the experiment, the probability of a call ending, the proportion of calls made to contacts from the contact list and their size.

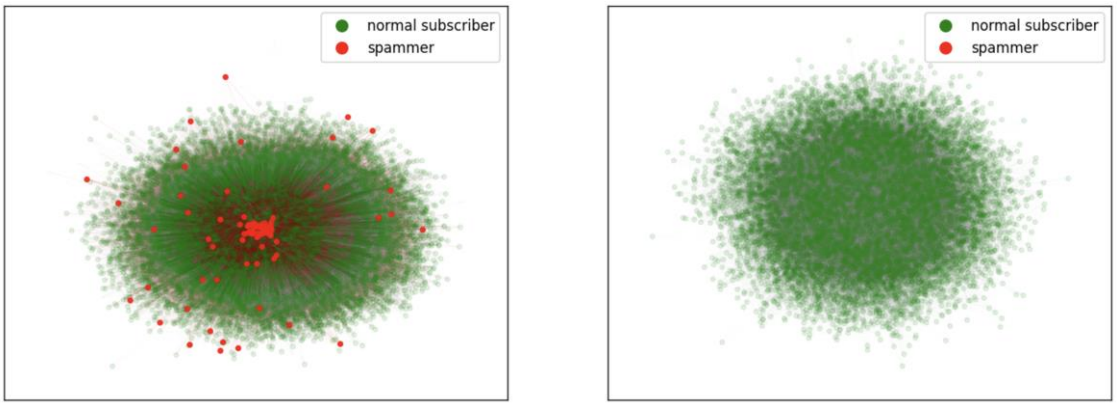
In the numerical experiments, various network properties of subscribers, or network nodes, were measured: node degree, local and global clustering coefficients, node centrality, average shortest path length between subscribers, PageRank [24], and the average number of calls per day [25, 26]. The degree of a node in a graph is defined as the number of edges connected to that node [25], indicating how many other nodes are directly connected to it. The clustering coefficient quantifies how connected a node's neighbours are in the graph [26]. This measure is expressed in two forms: the local clustering coefficient and the global clustering coefficient. Both provide insight into the degree of clustering in the network but at different scales. The local clustering coefficient of a node measures the proportion of

pairs of a node's neighbours that are directly connected to each other, relative to the total number of possible connections among those neighbours. In contrast, the global clustering coefficient measures the general tendency of the network to form closely related groups or clusters. The average shortest path length is defined as the average number of steps along the shortest paths between all possible pairs of nodes in the graph [25].

The employed model enables the analysis of subscriber behaviours and interactions through key network properties. By measuring network properties, we can identify patterns indicative of abnormal behaviour. This approach offers valuable insights for detecting spammers or subscribers with abnormal activity.

3 Detecting the presence of spammers within the network

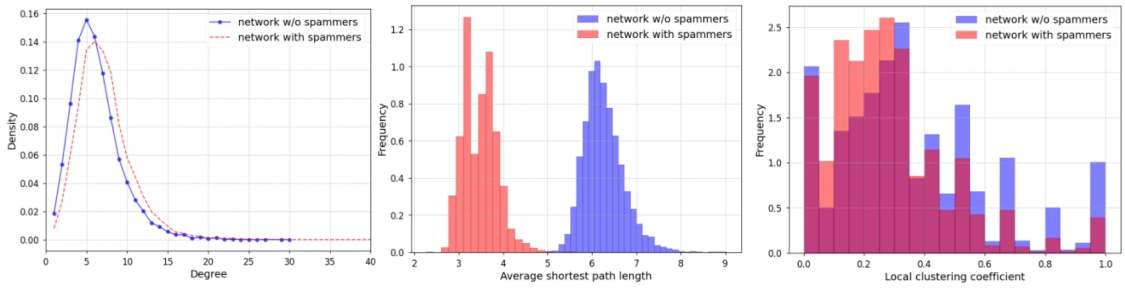
Detecting the presence of spammers in a dynamic telephone network involves analyzing how their behaviour influences key structural properties of the network. In this study, we simulated a dynamic telephone network for one day, consisting of 10,000 subscribers, with 1% (100) of them being spammers. Normal subscribers made 85% of calls to the people in their contact list, approximating real-world network conditions as described in [27]. Meanwhile, spammers called randomly. Based on the simulation data, two networks were created: one containing both normal subscribers and spammers (Fig 1.a), and the other consisting solely of normal subscribers (Fig 1.b).



a) *b)*

Fig. 1 Simulated telephone subscriber networks:
a) with spammers and normal subscribers, b) with normal subscribers

To detect the presence of malicious subscribers in the network, metrics such as node degree, global and local clustering coefficients, and average shortest path length were utilized. Based on the modeling data and the resulting networks, we constructed distributions for the analyzed metrics (Fig. 2) and computed their statistical characteristics (Table 1).



a) *b)* *c)*

Fig.2 Distributions of a) node degrees; b) average shortest path length;
c) local clustering coefficient. Blue color corresponds to the characteristics of a network without spammers, and red color corresponds to a network with spammers.

Table 1 – Characteristics of the studied networks

Network	Degree		Shortest path length		Clustering (local)		Clustering (global)
	mean	sd	mean	sd	mean	sd	
with spammers	8.12	27.06	3.48	0.39	0.28	0.20	0.02
without spammers	6.24	3.14	6.26	0.45	0.36	0.25	0.29
difference	1.88	-	-2.78	-	-0.08	-	-0.28
	23%	-	-80%	-	-28%	-	-1579%

The average degree is higher in the network with spammers (8.12) compared to the network without them (6.24). These users tend to connect to a large number of other nodes, often through actions like sending messages or making calls to numerous users, which increases their degree. A significantly higher average degree or the presence of nodes with abnormally high degrees may indicate the presence of spammers. However, not all of them may be active, and their behaviour can be difficult to distinguish from that of normal subscribers based solely on node degree. The average shortest path length is considerably shorter in the network containing spam entities (3.48) than in the network without them (6.26). By establishing numerous connections, these subscribers create shorter paths across the network, thereby reducing the overall shortest path length. This reduction suggests that the network is becoming more tightly connected, but in a potentially anomalous way, indicative of spam-like behaviour. The local clustering coefficient is slightly lower in the network with spammers (0.28) compared to the network without them (0.36). Such subscribers tend to connect to many unrelated nodes, resulting in a lower local clustering coefficient. This suggests that calls in the network are more random and less structured, resembling a pattern typical of abnormal activity. The global clustering coefficient is significantly lower in the network with spammers (0.02) compared to the network without them (0.29). They disrupt clustering by establishing connections across the network in a broad, random manner, which lowers the global clustering coefficient. A sharp decrease in this metric may serve as a strong indicator of abnormal users in a network, as it reflects a breakdown in the natural clustering and organization of the network.

Although node degree is the most straightforward metric to calculate, its usefulness is limited when spammers frequently change numbers. The shortest path length, while effectively distinguishing between networks with and without spammers, is computationally expensive when calculated for all pairs of subscribers. The local clustering coefficient is also sensitive to the presence of spammers but exhibits larger standard deviations. In contrast, the global clustering coefficient is more sensitive and computationally less expensive than the average shortest path length, making it the most suitable indicator for detecting users with abnormal behavour in a telephone subscriber network.

To test the sensitivity of the global clustering coefficient to the presence of spammers, we simulated their appearance in the network over time. The simulation was conducted with the same parameters as before for a 15-day period, with spam-like activity occurring from the fifth to the eleventh day. Using the obtained data, we calculated the global clustering coefficient for each day and plotted its dynamic behaviour over time (Fig. 3).

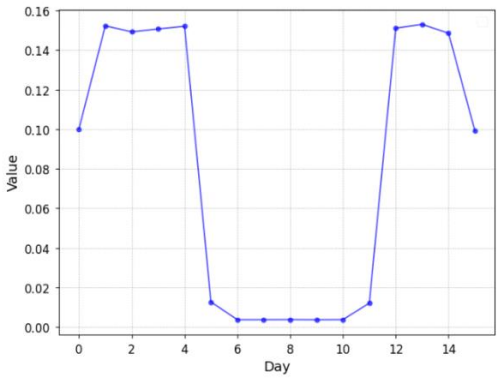


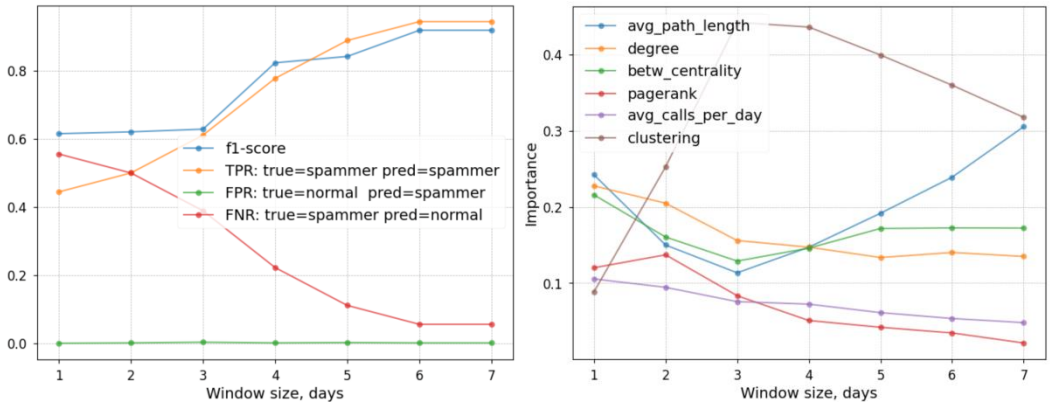
Fig.3 Dynamics of the global clustering coefficient with spammers (5-11 days) and without.

The dependence shown in Fig. 3 illustrates the high sensitivity of the clustering coefficient, with changes of more than an order of magnitude observed, to the presence of spammers in the dynamic

telephone subscriber network. Thus, this coefficient can serve as a clear indicator of abnormal activity within the network. In studies such as [9] and [28], PageRank and the local clustering coefficient have been employed to identify malicious users in systems like Skype and banking networks. These metrics are calculated individually for each user and utilized in classifier models to detect such users. In contrast, the global clustering coefficient can be used to assess the presence of spammer activity across the entire network, before applying classifiers to individual subscribers.

4 Subscriber classification and network properties measurement window size

To classify subscribers and assess the effect of the subscriber properties measurement window size, a numerical experiment was conducted with 10,000 subscribers, 1% of whom were spammers. Random Forest was selected as the classifier due to its minimal requirements for input data distributions, its tolerance for unnormalized or unprocessed data, and its ability to easily generate feature importance values. 80% of the subscribers were included in the training dataset, while the remaining 20% were used for testing. Given that the network of telephone subscribers is dynamic and evolves over time based on subscriber activity, the properties of subscribers also change accordingly. To evaluate the impact of the measurement period on network properties, we constructed seven classifiers, each trained with subscriber properties measured over varying time windows, starting from 1 day and gradually increasing the window size to 7 days. The results of subscriber classification and the importance of subscriber properties for spammer identification are presented in Fig. 4.



a) *b)*

Fig. 4 a) classifier accuracy indicators; b) importance of subscriber properties;

Fig. 4.a. presents the accuracy metrics of the classifier. The F1 score is depicted as a blue curve, the true positive rate (TPR) is shown as an orange curve, the false positive rate (FPR) is represented by a green curve, and the false negative rate (FNR) is indicated by a red curve. Fig. 4.b. illustrates the properties of subscribers and their relative importance within the classification model. The shortest path length (avg_path_length) is shown as a blue curve, the node degree (degree) is represented by an orange curve, node betweenness centrality (betw_centrality) is depicted as a green curve, PageRank is shown as a red curve, the average number of calls per day (avg_calls_per_day) is represented by a purple curve, and the local clustering coefficient (clustering) is shown as a brown curve.

The results of the classifier evaluation demonstrate that the accuracy of spammer detection reaches acceptable levels, with an F1 score greater than 0.80 when the subscriber properties calculation window is set to 4 days. Notably, the number of calls per day is a key distinguishing factor between spammers and normal subscribers; however, it is the least important feature for classification, with its significance diminishing as the window size increases. This can be explained by the fact that some spammers make the same number of calls as normal subscribers, and conversely, some normal subscribers may call excessively. In such cases, when classification based on call frequency is insufficient, network properties provide additional discriminatory power. As shown in Fig. 4b, the three most important properties for classification across all window sizes (1 to 7 days) are node degree, shortest path length, and betweenness centrality. However, for window sizes up to 3 days, these properties, along with the clustering coefficient, are not sufficient for effective spammer detection, as reflected in the lower F1 score and TPR. The F1 score and TPR reach values of 0.81 and 0.79, respectively, only when the window size is 4 days, with the importance of the average shortest path length increasing. As the

window size continues to grow, classifier performance improves, while the relative importance of the local clustering coefficient decreases and the average shortest path length becomes more significant.

5 Conclusion

The identification of telephone subscribers engaged in malicious activities, such as spamming or disturbing normal users, is an important issue in network security. This study explores the potential for detecting the presence of spammers within a telephone subscriber network and classifying individual subscribers. A distinguishing feature of this research is the use of network characteristics as subscriber properties. The study employs a telephone network model to investigate these characteristics. It was found that, to detect the presence of spammers without classifying individual subscribers, the global clustering coefficient is a sufficient metric. This indicator demonstrated high sensitivity with its value decreasing significantly from 0.29 to 0.02 following the emergence of spammers in the network. Furthermore, the study examines the use of subscriber network properties for spammer identification using the Random Forest machine learning model. The importance of network properties was evaluated over seven measurement intervals, ranging from 1 to 7 days. Key properties identified as most important for classification included the local clustering coefficient, average shortest path length, node degree, and betweenness centrality. The study found that, with a measurement window of 4 days, the classifier achieved F1 score of 80%, along with a high spammer detection rate (TPR). These findings suggest that network characteristics significantly enhance model accuracy, although a certain period of observation is required to effectively distinguish between spammers and normal subscribers. Consequently, network characteristics are valuable for identifying dormant or evasive subscribers – those who frequently change numbers, avoid detection through call frequency or blocklists, or remain inactive for extended periods.

REFERENCES

1. N. Davey, S. Field, R. Frank, P. Barson, and G. McAskey, “The detection of fraud in mobile phone networks”, Neural Network World, vol. 6, no. 4, pp. 477–484, 1996.
2. H. Li et al., “A Machine Learning Approach To Prevent Malicious Calls Over Telephony Networks”, in 2018 IEEE Symposium on Security and Privacy (SP), May 2018, pp. 53–69. DOI: 10.1109/SP.2018.00034. (Last accessed: 15.11.2024).
3. “FCC fines illegal robocalling company a record-breaking $300 MILLION after it made more than five billion calls to more than 500M phone numbers in a three-month span | Daily Mail Online.” Accessed: Aug. 16, 2024. URL: https://www.dailymail.co.uk/news/article-12371697/FCC-fines-illegal-robocalling-company-record-breaking-300-MILLION-five-billion-calls-500M-phone-numbers-three-month-span.html (Last accessed: 15.11.2024).
4. N. Jiang et al., “Isolating and analyzing fraud activities in a large cellular network via voice call graph analysis”, in Proceedings of the 10th international conference on Mobile systems, applications, and services, Low Wood Bay Lake District UK: ACM, Jun. 2012, pp. 253–266. DOI: 10.1145/2307636.2307660. (Last accessed: 15.11.2024).
5. H. Tu, A. Doupe, Z. Zhao, and G.-J. Ahn, “SoK: Everyone Hates Robocalls: A Survey of Techniques Against Telephone Spam”, in 2016 IEEE Symposium on Security and Privacy (SP), San Jose, CA: IEEE, May 2016, pp. 320–338. DOI: 10.1109/SP.2016.27. (Last accessed: 15.11.2024).
6. P. Patankar, G. Nam, G. Kesidis, and C. R. Das, “Exploring Anti-Spam Models in Large Scale VoIP Systems”, in 28th IEEE International Conference on Distributed Computing Systems (ICDCS 2008), 17-20 June 2008, Beijing, China, IEEE Computer Society, 2008, pp. 85–92. DOI: 10.1109/ICDCS.2008.71. (Last accessed: 15.11.2024).
7. F. Wang, Y. Mo, and B. Huang, “P2P-AVS: P2P Based Cooperative VoIP Spam Filtering”, in Proceedings of the 2007 IEEE Wireless Communications and Networking Conference, USA: IEEE Computer Society, 2007, pp. 3547–3552. DOI: 10.1109/WCNC.2007.650. (Last accessed: 15.11.2024).
8. N. Jiang, Y. Jin, A. Skudlark, and Z.-L. Zhang, “Greystar: fast and accurate detection of SMS spam numbers in large cellular networks using grey phone space”, in Proceedings of the 22nd USENIX Conference on Security, in SEC’13. USA: USENIX Association, 2013, pp. 1–16. DOI: 10.5555/2534766.2534768. (Last accessed: 15.11.2024).

9. A. Leontjeva, M. Goldszmidt, Y. Xie, F. Yu, and M. Abadi, “Early security classification of skype users via machine learning”, in Proceedings of the 2013 ACM workshop on Artificial intelligence and security, Berlin Germany: ACM, Nov. 2013, pp. 35–44. DOI: 10.1145/2517312.2517322. (Last accessed: 15.11.2024).
10. Y. Rebahi, D. Sisalem, and T. Magedanz, “SIP Spam Detection”, in International Conference on Digital Telecommunications (ICDT’06), Cote d’Azur, France: IEEE, 2006, pp. 68–68. DOI: 10.1109/ICDT.2006.69. (Last accessed: 15.11.2024).
11. N. Miramirkhani, O. Starov, and N. Nikiforakis, “Dial One for Scam: A Large-Scale Analysis of Technical Support Scams”, in Proceedings 2017 Network and Distributed System Security Symposium, San Diego, CA: Internet Society, 2017. DOI: 10.14722/ndss.2017.23163. (Last accessed: 15.11.2024).
12. S. Subudhi and S. Panigrahi, “Use of Possibilistic Fuzzy C-means Clustering for Telecom Fraud Detection”, in Computational Intelligence in Data Mining, vol. 556, H. S. Behera and D. P. Mohapatra, Eds., in Advances in Intelligent Systems and Computing, vol. 556. , Singapore: Springer Singapore, 2017, pp. 633–641. DOI: 10.1007/978-981-10-3874-7_60. (Last accessed: 15.11.2024).
13. S. Subudhi and S. Panigrahi, “Quarter-Sphere Support Vector Machine for Fraud Detection in Mobile Telecommunication Networks”, Procedia Computer Science, vol. 48, pp. 353–359, 2015, DOI: 10.1016/j.procs.2015.04.193. (Last accessed: 15.11.2024).
14. R. Zhang and A. Gurtov, “Collaborative Reputation-based Voice Spam Filtering”, in 2009 20th International Workshop on Database and Expert Systems Application, Linz, Austria: IEEE, 2009, pp. 33–37. DOI: 10.1109/DEXA.2009.95. (Last accessed: 15.11.2024).
15. D. Ucci, R. Perdisci, J. Lee, and M. Ahamad, “Building a Collaborative Phone Blacklisting System with Local Differential Privacy”, Jun. 16, 2020, arXiv: arXiv:2006.09287. URL: http://arxiv.org/abs/2006.09287 (Last accessed: 15.11.2024).
16. J. Daka and M. Nyirenda, “Smart Mobile Telecommunication Network Fraud Detection System Using Call Traffic Pattern Analysis and Artificial Neural Network”, vol. 12, pp. 43–50, Apr. 2023, DOI: 10.5923/j.ajis.20221202.01. (Last accessed: 15.11.2024).
17. Q. Zhao, K. Chen, T. Li, Y. Yang, and X. Wang, “Detecting telecommunication fraud by understanding the contents of a call”, Cybersecur, vol. 1, no. 1, p. 8, Dec. 2018, DOI: 10.1186/s42400-018-0008-5. (Last accessed: 15.11.2024).
18. Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria, B. O. Akinyemi, O. H. Odukoya, M. L. Sanni, G. Sewagnon, and G. A. Aderounmu, “Performance Evaluation of Machine Learning based Robocalls Detection Models in Telephony Networks” IJCNIS, vol. 14, no. 6, pp. 37–53, Dec. 2022, DOI: 10.5815/ijcnis.2022.06.04. (Last accessed: 15.11.2024).
19. M. Fire, G. Katz, and Y. Elovici, “Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies”, Human journal, vol. 1, no. 1, pp. 26–39, 2012.
20. C. Chaparro and W. Eberle, “Detecting Anomalies in Mobile Telecommunication Networks Using a Graph Based Approach”. URL: https://cdn.aaai.org/ocs/10377/10377-46053-1-PB.pdf. (Last accessed: 15.11.2024).
21. X. Hu, H. Chen, H. Chen, X. Li, J. Zhang, and S. Liu, “Mining Mobile Network Fraudsters with Augmented Graph Neural Networks”, Entropy, vol. 25, no. 1, p. 150, Jan. 2023, DOI: 10.3390/e25010150. (Last accessed: 15.11.2024).
22. S. Ji, J. Li, Q. Yuan, and J. Lu, “Multi-Range Gated Graph Neural Network for Telecommunication Fraud Detection”, in 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom: IEEE, Jul. 2020, pp. 1–6. DOI: 10.1109/IJCNN48605.2020.9207589. (Last accessed: 15.11.2024).
23. V. Danilevskiy and V. Yanovsky, “Statistical properties of telephone communication network”, arXiv preprint arXiv:2004.03172, 2020. (Last accessed: 15.11.2024).
24. S. Brin and L. Page, “The anatomy of a large-scale hypertextual Web search engine”, Computer Networks and ISDN Systems, vol. 30, no. 1, pp. 107–117, 1998, DOI: https://doi.org/10.1016/S0169-7552(98)00110-X. (Last accessed: 15.11.2024).
25. M. Newman, Networks, vol. 1. Oxford University Press, 2018. DOI: 10.1093/oso/9780198805090.001.0001. (Last accessed: 15.11.2024).

26. D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks”, Nature, vol. 393, no. 6684, pp. 440–442, Jun. 1998, DOI: 10.1038/30918. (Last accessed: 15.11.2024).
27. M. Danilevskyi, V. Yanovsky, and O. Matsiy, “Modeling and Analysis of a Dynamic Network of Telephone Subscribers Considering the Degree of Connectivity by Means of Contact Lists (Unpublished article)”, Bulletin of V.N. Karazin Kharkiv National University, series «Mathematical modeling. Information technology. Automated control systems».
28. C. Azarm, E. Acar, and M. van Zeelt, “On the Potential of Network-Based Features for Fraud Detection”, Feb. 19, 2024, arXiv: arXiv:2402.09495. URL: http://arxiv.org/abs/2402.09495 (Last accessed: 15.11.2024).

Данілевський Михайло Вікторович *аспірант; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022*
e-mail: m.danilevskiy@gmail.com
https://orcid.org/0009-0000-0030-2218

Яновський Володимир Володимирович *доктор фізико-математичних наук, професор, професор кафедри штучного інтелекту та програмного забезпечення Харківський національний університет імені В. Н. Каразіна, майдан Свободи 4, Харків-22, Україна, 61022 Завідувач теоретичним відділом, інститут монокристалів НАН України, пр.Науки 60, Харків, Україна, 61001*
e-mail: yanovsky@isc.kharkov.ua
https://orcid.org/0000-0003-0461-749X

Виявлення телефонних абонентів із аномальною поведінкою за допомогою аналізу властивостей мережі

Актуальність. Використання мережі телефонних абонентів з метою шахрайства, продажу товарів та послуг, спаму призводить до щорічних фінансових втрат у мільярди доларів у всьому світі. Проблема вивчалася ще 1996р. і до сьогодні є актуальною, незважаючи на безліч способів протидії та захисту. Традиційні методи, такі як списки блокування та моніторинг частоти викликів, часто неефективні проти спамерів, які обходять ці системи, змінюючи моделі поведінки.

Мета. Метою роботи є вивчення використання мережевих властивостей абонентів, таких як коефіцієнти кластеризації, центральність та середня довжина найкоротшого шляху, як критерії для виявлення абонентів з аномальною поведінкою в динамічній телефонній мережі.

Методи дослідження. Моделювання та чисельний експеримент.

Результати. Дослідження показує, що глобальний коефіцієнт кластеризації є чутливою мірою виявлення присутності спамерів в мережі. Його значення знижується в кілька разів при появі спамерів. При класифікації абонентів на звичайний та спамер за допомогою моделі Random Forest, найважливішими властивостями є локальний коефіцієнт кластеризації, середня довжина найкоротшого шляху, ступінь та центральність абонента в мережі. За даними моделювання мережі телефонних абонентів було виявлено, що при розмірі вікна вимірювання у 4 дні показник точності класифікатора (F1 score) та точності виявлення спамерів (TPR) досягає значень 80%.

Висновки. Використання мережевих характеристик абонентів позитивно впливає на точність виявлення абонентів з аномальною поведінкою, але при цьому вимагає часу, щоб спамери та звичайні абоненти стали помітними за мережевими характеристиками.
Ключові слова: *надокучливі телефонні дзвінки, граф викликів, телефонний спам, телефонна мережа, логнормальний розподіл, розподіл ступенів, коефіцієнт кластеризації, середня довжина найкоротшого шляху*.

Надійшла у першій редакції 02.09.2024, в останній - 09.10.2024.
The first version has been received on 02.09.2024, the final version - on 09.10.2024.