ДЕТЕКТИРОВАНИЕ БОТОВ В СОЦИАЛЬНЫХ СЕТЯХ

С.В. Брызгалов, Я.Р. Мустакимова, А.Н. Рабчевский
Пермский государственный национальный исследовательский университет

Аннотация. В статье рассматривается проблема применения ботов в социальных сетях, которые все чаще используются как средства для осуществления информационных атак. Целью данной работы является анализ текущего состояния методов выявления и противодействия ботам, а также разработка программы для их детектирования. С помощью разработанной программы был проведен анализ нескольких информационных атак, проведенных в социальной сети ВКонтакте. В результате проведенных исследований был сделан вывод о том, что мониторинг активности ботов может служить важным индикатором для идентификации новых информационных атак.
Ключевые слова: *информационные атаки, социальные сети, боты, детектирование ботов.*

Введение

В последнее время социальные сети стали неотъемлемой частью нашей повседневной жизни. Миллионы людей по всему миру пользуются ими для общения, обмена информацией и поиска развлекательного контента. С увеличением аудитории социальных сетей возникла потребность в автоматизации процесса управления контентом, что привело к популярности ботов.

В общем случае под ботом понимается программа, автоматически выполняющая заранее настроенные повторяющиеся задачи [1]. Боты обычно либо имитируют поведение пользователя, либо полностью заменяют его. В социальных сетях боты могут использоваться для различных целей. Например, с помощью ботов можно эффективно распространять информацию. Боты могут автоматически публиковать новости и объявления. Многие компании используют ботов для проведения рекламных кампаний, анализа пользовательских данных и формирования рекомендаций на основе поведения пользователей. Боты могут вести беседы с пользователями, предоставлять информацию о каких-либо продуктах или услугах.

Но также боты могут представлять собой серьезную угрозу. Некоторые боты специально создаются злоумышленниками для распространения фейковых новостей и манипуляции общественным мнением. Например, во время пандемии коронавируса боты в Twitter подвергали сомнению правдивость информации, распространяемой через официальные источники, и массово рассылали ссылки на недостоверные источники информации. Также боты распространяли посты с критикой мер, принятых для сдерживания пандемии [2]. Еще боты в социальных сетях могут использоваться

для шантажа или кражи личной информации [3]. В работе [4] отмечается, что вредоносные боты являются одним из основных инструментов проведения атак в социальных сетях.

Но несмотря на существенный прогресс в развитии методик выявления и противодействия ботам, методы определения характеристик ботов на данный момент развиты недостаточно. Боты не просто используются в информационных атаках, но и постоянно эволюционируют, что в свою очередь влияет на возможности их детектирования. Поэтому важно разрабатывать и совершенствовать средства, позволяющие отличить настоящих пользователей от ботов.

Обзор литературы

Согласно оценке компании Arkose Labs, 73% всего интернет-трафика в настоящее время составляют и распространяют боты [5]. Конечно, некоторые боты выполняют полезные функции, но большинство созданы для вредоносных целей.

Боты способны выполнять тысячи задач одновременно, что позволяет злоумышленникам не только экономить время, но и значительно увеличивать масштаб своих операций. С помощью ботов можно маскировать свое истинное местоположение и личность, что усложняет работу правоохранительным органам по выявлению и задержанию преступников. С помощью ботов можно манипулировать эмоциями и поведением реальных пользователей социальных сетей. В исследовании [6] говорится, что даже незначительное количество ботов в онлайн-среде может существенно изменить восприятие и взгляды реальных пользователей. Это происходит благодаря тому, что боты способны создавать иллюзию массовой поддержки определенных мнений и позиций, что влияет на общественное мнение. Также работа ботов может способствовать формированию фильтров, что тоже является серьезной проблемой. Алгоритмы социальных сетей, предназначенные для персонализации контента, часто создают фильтр, в результате работы которого пользователи сталкиваются преимущественно с теми мнениями, которые совпадают с их собственными. Например, если пользователь активно взаимодействует с определенным видом контента или поддерживает определённые взгляды, то боты начнут подбирать и показывать ему больше информации в этом же ключе, игнорируя альтернативные точки зрения. Это создает искаженное восприятие реальности и ограничивает критическое мышление человека. По всем этим причинам злоумышленники используют ботов при проведении информационных атак.

Для детектирования ботов можно использовать следующие методы:

1. Метод «живой силой». В данном случае поиском ботов занимаются модераторы. Они ищут ботов на основе жалоб от пользователей. Модераторы исследуют профиль пользователя: обращают внимание на наличие неполной информации, на отсутствие фотографии, на одинаковые или общие темы в публикациях и комментариях. Также модераторы отслеживают

необычные шаблоны поведения пользователей. Например, слишком высокую активность пользователя или работу пользователя в круглосуточном режиме. Однако модератор может потерять концентрацию внимания в течение нескольких минут рутинной работы, в следствии чего он может с легкостью не увидеть какой-то критически важный параметр или наоборот, добавить что-то от себя, чем уменьшит точность предсказания. Также при увеличении количества пользователей социальных сетей необходимо нанимать большее число модераторов.

2. Анализ структуры социального графа. Данный метод заключается в построении социального графа и его дальнейшем анализе. Для этого используются различные алгоритмы поиска компонент сильной связности [7]. При изучении топологии графа необходимо обратить внимание на узлы с необычными характеристиками. Например, боты могут иметь значительное количество подписчиков и подписок. Или если аккаунт имеет множество связей с другими пользователями, но не взаимодействует с ними, то это может быть подозрительным. Также боты могут образовывать отдельные кластеры или взаимодействовать лишь с определёнными группами, отличающимися от большинства пользователей. К недостаткам данного метода можно отнести большие временные затраты на построение графа, а также низкую точность метода, потому что итоговые выводы формирует человек, который может ошибиться.

3. Метод машинного обучения. Данный метод позволяет свести к минимуму человеческий фактор и субъективизм, потому что алгоритмы опираются на данные и заранее заданные критерии, что снижает вероятность ошибок, связанных с персональным мнением или предвзятостью модератора.

Метод, использующий машинное обучение, включает в себя два основных подхода. Первый заключается в анализе контента, который публикует пользователь [8]. Здесь ключевым моментом является работа с текстами сообщений, изображениями и видео, которые могут содержать маркеры, указывающие на автоматизированное поведение. Этот подход особенно эффективен для социальных сетей, где контент имеет однородный характер, поскольку это упрощает процесс классификации и выявления аномалий. Второй подход основывается на изучении метаданных аккаунта пользователя. В данном случае анализируются такие данные, как имя владельца аккаунта, дата его создания, объем размещенных материалов, количество друзей и подписок и т.д. Одной из ключевых проблем данного подхода является выбор и извлечение релевантных признаков, поскольку для достижения высокой точности анализа важно правильно определить, какие характеристики будут наиболее информативными.

Конечно и у этого метода есть свои недостатки. Например, алгоритмы часто разрабатываются с учетом особенностей конкретных социальных сетей, что затрудняет использование одних и тех же программных решений для разных социальных платформ. Также возникает сложность со сбором качественных данных для обучения, так как часто используются не-

большие или плохо размеченные наборы данных. Тем не менее алгоритмы машинного обучения могут анализировать и обрабатывать огромные объемы данных непрерывно за небольшой промежуток времени, что позволяет быстро выявлять подозрительное поведение и обнаруживать ботов.

Программа для детектирования ботов

Целью нашей работы является создание программы, позволяющей выявлять ботов в социальной сети ВКонтакте. ВКонтакте – это одна из самых популярных социальных сетей в России и странах СНГ, обеспечивающая доступ к широкой и разнообразной аудитории. Пользователи сети могут легко делиться публикациями, фотографиями, видео и другими материалами, что способствует быстрому распространению информации. Платформа поддерживает различные форматы контента – текст, изображения, видео, аудио – позволяя обмениваться информацией в удобном для пользователей виде. Возможности комментирования, лайков и репостов позволяют пользователям взаимодействовать с контентом, что также способствует распространению информации. Именно поэтому эту социальную сеть злоумышленники очень часто используют в своих целях.

Рассматривая методы детектирования ботов, мы решили остановиться на методе машинного обучения, как наиболее простом и эффективном решении для анализа большого количества пользователей.

Просмотрев аккаунты пользователей ВКонтакте, мы выделили параметры, которые будут поданы на вход нейронной сети. В социальной сети ВКонтакте есть два вида профилей пользователей – открытые и закрытые. У открытых профилей доступно больше данных для анализа, чем у закрытых. У закрытых профилей возможно считать только те параметры, которые в таблице 1 выделены серым цветом, в то время как у открытых можно считать все представленные в таблице параметры.

Таблица

Параметры аккаунтов

Наличие галочки	Кол-во подписок	Ср. кол-во дней между постами	Ср. кол-во просмотров на постах	Кол-во видео
Есть ли имя в словаре имен	Дней с последнего захода	Отношение постов и репостов	Кол-во подписчиков	Кол-во аудио
Есть ли никнейм	Возраст аккаунта в днях	Кол-во постов с удаленными	Кол-во подписчиков в VK Клипы	Кол-во подарков
Полнота заполнения профиля	Кол-во постов на данный момент	Кол-во постов в день за последние 100 постов	Кол-во фото профиля	Кол-во групп
Кол-во друзей	Кол-во постов в день за время жизни аккаунта	Ср. кол-во комментариев на постах	Кол-во указаний на фото	Кол-во городов в группах
Кол-во «Интересных страниц»		Ср. кол-во лайков на постах	Кол-во фотоальбомов	Кол-во групп без аватарок

Разделение профилей на открытые и закрытые производится для более достоверного анализа. В связи с этим были созданы две разные нейросетевые модели – для поиска ботов среди открытых профилей и для поиска ботов среди закрытых профилей. Обе нейронные сети представляют собой многослойный персептрон с одним скрытым слоем и одним выходным нейроном, который дает нам число от 0 до 1, которое обозначает вероятность того, что данный аккаунт – бот.

В ходе разработки нейросетевых моделей была изучена значимость всех входных параметров. Самым низким значением обладает параметр «наличие галочки», максимальное значение имеет параметр «возраст аккаунта». В результате анализа стало ясно, что каждый параметр влияет на конечный результат, и избавление даже от самых незначительных может сказаться на точности.

Набор данных для обучения собирался вручную. Аккаунты ботов были получены благодаря сайту Ботнадзор [9]. В качестве обычных пользователей были взяты аккаунты из групп различных школ, университетов, подслушано разных городов.

Проверка полученных нейросетевых моделей на тестовых данных показала следующие результаты: для открытых профилей – 95% точности выявления ботов, для закрытых – 86%.

После успешного тестирования нейросетевых моделей была написана программа «Детектор ботов» [10]. Она классифицирует профили пользователей ВКонтакте на две категории – «Боты» и «Не боты». Работу программы можно представить в виде следующей последовательности шагов:

1. На первом шаге осуществляется формирование текстового файла с ссылками на профили пользователей. Таким образом, любые профили ВКонтакте можно проверить, записав их в файл и передав его программе.

2. На втором шаге проводится разделение всех профилей на три группы: «Открытые», «Закрытые» и «Удаленные». Принцип разделения профилей по категориям, следующий: программа через VK API [11] проверяет каждый отдельный профиль и смотрит к какой группе его отнести. Все профили из категории «Удаленные» определяются как боты. Проанализировать удаленный профиль и получить какие-либо данные относительно его владельца нельзя, а вероятность того, что профиль был удален самим пользователем значительно ниже, чем вероятность того, что модераторы ВКонтакте заметили подозрительную активность и заблокировали бота.

3. На третьем шаге происходит сбор данных о пользователях из групп «Открытые» и «Закрытые». Все данные, необходимые для определения параметров из таблицы 1, извлекаются с помощью VK API. Следует отметить, что программа собирает только данные профиля, не затрагивая публикации пользователя, кроме времени этих самых публикаций. Это сделано для оптимизации сбора данных, основываясь на следующей логи-

ке: если модель нейронной сети для открытых пользователей уже может классифицировать профили с 95% точностью, то добавление двух-трех процентов к точности от анализа публикаций не оправдают временные затраты на их обработку. А к закрытым профилям такое решение и вовсе не применимо, так как невозможно просмотреть публикации закрытых профилей.

Чтобы облегчить взаимодействие с VK API, были созданы специальные процедуры. Но эти процедуры накладывают некоторое ограничение на работу программы: списки профилей обрабатываются наборами размером не более 25 профилей в каждом. Данное решение обеспечивает некоторую степень сохранности от сбоев в работе программы. Если произойдет сбой, то потеряется только информация о последних 25 профилях. Информация о профилях, проанализированных ранее, сохранится.

4. На четвертом шаге осуществляется нейросетевой анализ. После сбора данных проводится их нормализация для того, чтобы ни один из параметров не был более важным просто потому, что он измеряется в десятках тысяч, а не в единицах. Все параметры должны быть равнозначными в возможностях своего влияния на итоговый результат. Параметры в данных проходят избавление от выбросов и нормализуются до значений от 0 до 1 с помощью метода скорректированного интервала, после чего подаются на вход нейросетевым моделям. На выходе мы получаем вероятность того, что проверяемый аккаунт является ботом. Для облегчения дальнейшей работы с этой информацией было принято решение считать все аккаунты, вероятность которых ниже 0.5 – «Не ботами», а те, у которых выше 0.5 – «Ботами».

Результаты

С помощью разработанной программы мы решили проверить как используются боты для проведения информационных атак в социальных сетях.

Мы проанализировали два информационных повода, по которым проводились информационные атаки в социальной сети ВКонтакте: «Дворец Путина» и «Специальная военная операция». Результаты показали, что для «Дворца Путина» в среднем 24% постов создаются ботами, тогда как для «Специальной военной операции» этот показатель значительно выше – 67%.

После чего нам стало интересно посмотреть, нет ли какой закономерности в том, когда используются боты. Мы предполагали обнаружить два теоретических шаблона (рис. 1). Первый – это постепенное уменьшение процента постов ботов со временем, когда реальные пользователи подхватывают сообщение и начинают распространять его сами. То есть людям становится интересна определенная информация и боты, как инструмент распространения, отходят на задний план. Нам показалось, что

это будет выглядеть более естественно с точки зрения распространения информации. Второй – это постепенное увеличение процента постов ботов от времени. В этом случае люди не проявляют интереса к определенной информации, не распространяют её самостоятельно, и эту задачу берут на себя боты. Ведь конечная цель информационной атаки – это довести нужную злоумышленнику информацию до обычных людей, а чем больше постов с этой информацией, тем выше шанс, что пользователи её увидят.

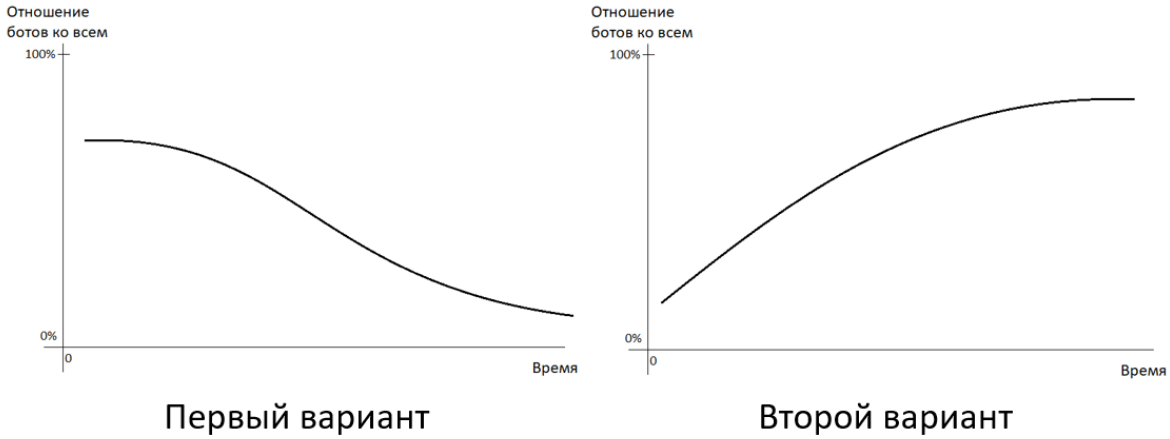

Рис. 1. Теоретические шаблоны

На практике мы получили следующие результаты. Во-первых, мы выявили случаи (рис. 2), когда пост сначала создавался и публиковался человеком, затем его начали распространять обычные пользователи, а позже к этому процессу подключились и боты. В таких случаях можно заметить, что соотношение постов, генерируемых ботами, к общему числу постов остается на постоянном уровне на протяжении всего времени.

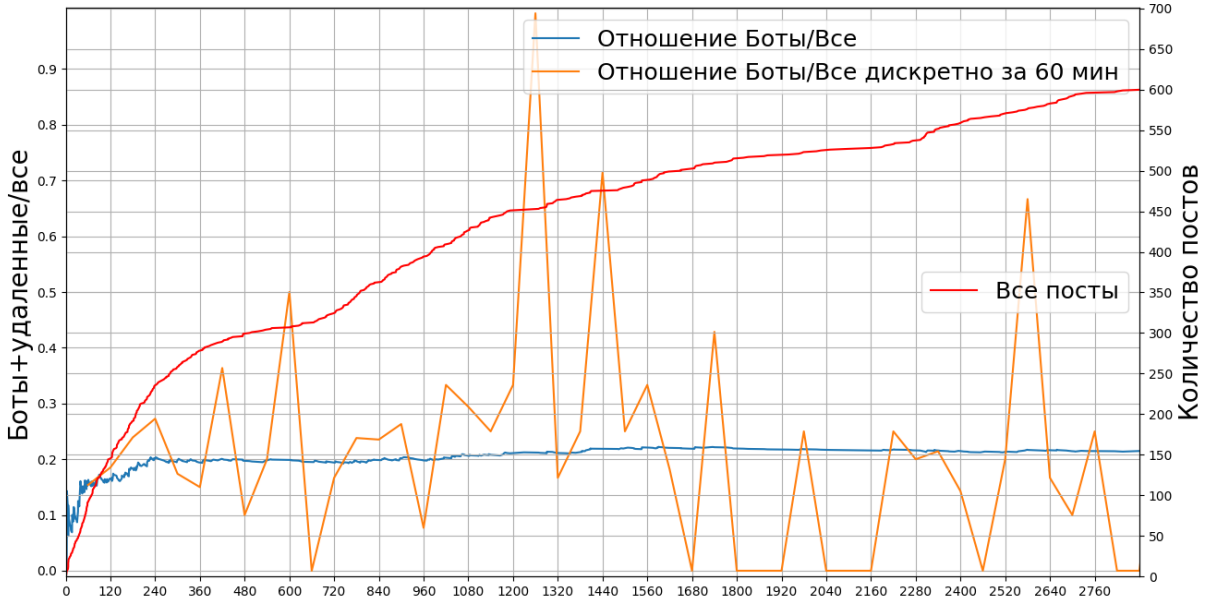

Рис. 2. Практический шаблон 1

Во-вторых, есть случаи (рис. 3), когда первые несколько постов публикуются ботами, а затем эти посты начинают распространять люди совместно с ботами.

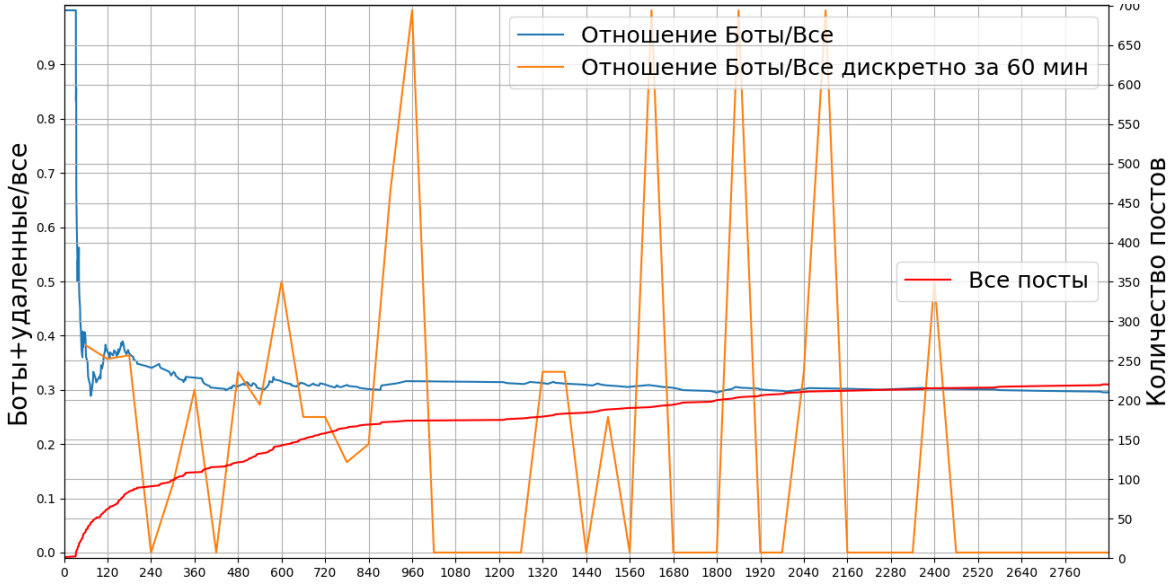

Рис. 3. Практический шаблон 2

По сути, оба этих случая подходят под первый теоретический шаблон, но вместо того чтобы снижать активность, боты продолжают свою работу.

Также мы выявили случаи (рис. 4), когда посты распространяют исключительно боты. Например, один и тот же пост боты размещали в нескольких группах в течении пары месяцев. Такие случаи вписываются во второй теоретический шаблон, когда сообщения преимущественно распространяются ботами.

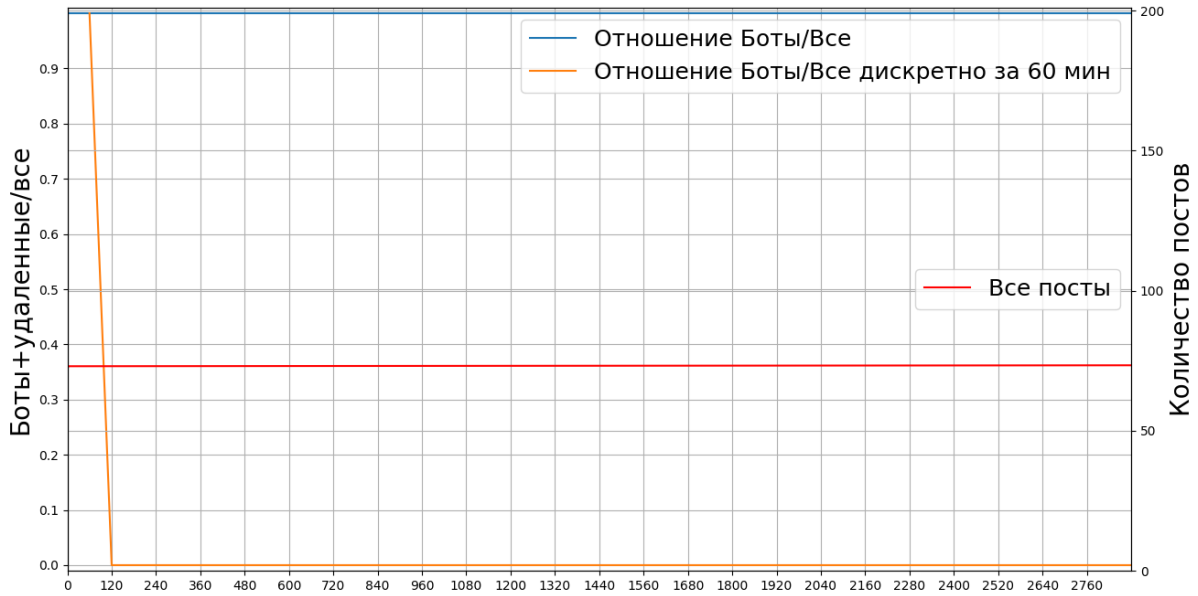

Рис. 4. Практический шаблон 3

Заключение

Проведенные исследования показывают, что злоумышленники активно используют ботов для проведения информационных атак в социальных сетях. Эти автоматизированные программы позволяют им эффективно распространять ложную информацию и манипулировать общественным мнением, что создает серьезные риски для пользователей и общества в целом. Учитывая этот факт, можно сделать вывод, что наблюдение за активностью таких ботов и их поведением в сети может стать важным индикатором для идентификации новых информационных атак.

Список литературы

1. Что такое боты – определение и описание [Электронный ресурс] URL: https://www.kaspersky.ru/resource-center/definitions/what-are-bots (дата обращения: 10.12.2024)
2. Suarez-Lledo V., Alvarez-Galvez J. Assessing the Role of Social Bots During the COVID-19 Pandemic: Infodemic, Disagreement, and Criticism // Journal of Medical Internet Research. 2022. Vol.24. №8. DOI: 10.2196/36085 EDN: IGSQBN
3. Об опасности использования Telegram-ботов [Электронный ресурс] URL: https://www.gazeta.ru/social/news/2023/11/29/21812269.shtml (дата обращения: 10.12.2024)
4. Коломеец М.В., Чечулин А.А. Метрики вредоносных социальных ботов // Труды учебных заведений связи. 2023. Т.9. №1. С.94-104. DOI: 10.31854/1813-324X-2023-9-1-94-104 EDN: HEFHFR
5. Breaking (Bad) Bots: Bot Abuse Analysis and Other Fraud Benchmarks [Электронный ресурс] URL: https://www.arkoselabs.com/wp-content/uploads/Breaking-Bad-Bots-Bot-Abuse-Analysis-and-Other-Fraud-Benchmarks.pdf (дата обращения: 10.12.2024)
6. Aldayel A., Magdy W. Characterizing the role of bots' in polarized stance on social media // Social Network Analysis and Mining. 2022. V.12, №30. DOI: 10.1007/s13278-022-00858-z EDN: TNWDBG
7. Danezis G., Mittal P. SybilInfer: Detecting Sybil Nodes using Social Networks // Conference: Proceedings of the Network and Distributed System Security Symposium, NDSS 2009, San Diego, California, USA.
8. Zhengab X., Zengab Zh., Chenc Zh., Yuab Y., Rong Ch. Detecting spammers on social networks // Neurocomputing. 2015. V.159. P.27-34. DOI: 10.1016/j.neucom.2015.02.047
9. Ботнадзор [Электронный ресурс] URL: https://botnadzor.org/ (дата обращения: 10.12.2024)
10. Программа «Детектор ботов» свидетельство о регистрации программы для ЭВМ регистрационный № 2024617353 от 01.04.2024.
11. API VK для разработчиков [Электронный ресурс] URL: https://dev.vk.com/ru/reference (дата обращения: 10.12.2024)

BOT DETECTION IN SOCIAL NETWORKS

S.V. Bryzgalov, Y.R. Mustakimova, A.N. Rabchevsky
Perm State University

Abstract. The article deals with the problem of using bots in social networks, which are increasingly being used as a means for information attacks. The purpose of this paper is to analyze the current state of methods for detecting and counteracting bots, as well as to develop a program for their detection. With the help of the developed program was analyzed several information attacks conducted in the social network VKontakte. As a result of the research it was concluded that monitoring of bot activity can serve as an important indicator for the identification of new information attacks.
Keywords: *information attacks, social networks, bots, bot detection*.