УДК 004

Исследование и анализ методов распознавания виртуальных посетителей страницы в социальной сети

Кукушкин Алексей Михайлович
Волжский политехнический институт (филиал) «Волгоградский государственный технический университет»
студент

Лясин Дмитрий Николаевич
Волжский политехнический институт (филиал) «Волгоградский государственный технический университет»
к.т.н., доцент

Рыбанов Александр Александрович
Волжский политехнический институт (филиал) «Волгоградский государственный технический университет»
к.т.н., доцент, зав. кафедрой информатика и технология программирования

Аннотация
В данной статье рассматриваются вопросы распознавания виртуальных посетителей страницы, бот-программ, создаваемых под видом реальных пользователей социальной сетей и имитирующих активность, в целях получения выгоды для создателя. В статье описываются типы бот-программ, способы их возможного использования, важность их выявления, а также рассмотрены превентивные методы защиты от применения бот-программ и выявлены критерии, позволяющие идентифицировать в социальных сетях бот-профили.
Ключевые слова: Социальные сети, бот, способы распознавания ботов.

Research and analysis of the methods of recognition of virtual visitors of the page in the social network

Kukushkin Alexey Mikhailovich
Volzhsky Polytechnical Institute (branch) of Volgograd State Technical University
student

Lyasin Dmitry Nikolaevich
Volzhsky Polytechnical Institute (branch) of Volgograd State Technical University
Candidate of Engineering Sciences, associate professor

Rybanov Aleksandr Aleksandrovich
Volzhsky Polytechnical Institute (branch) of Volgograd State Technical University

Candidate of Engineering Sciences, associate professor, Head of the Department of Informatics and programming technology

Abstract
This article discusses the recognition of virtual visitors to the page, bot programs created under the guise of real users of social networks and simulating activity, in order to obtain benefits for the Creator. The article describes the types of bot programs, ways of their possible use, the importance of their detection, as well as preventive methods of protection against the use of bot programs and the criteria for identifying bot profiles in social networks.
Keywords: Social networks, bot, virtual users, methods of bots recognition.

Социальные сети динамично развивающаяся сфера с постоянным притоком новых пользователей, предоставляющих на добровольной основе свои персональные и конфиденциальные данные. Ежегодно новых пользователей социальных сетей в мире увеличивается на 13 %. Среднестатистический пользователь проводит в соц. сети более 2 часов в сутки, а молодёжь (16–24 лет) - три часа, на основании только этих цифр видны возможности для маркетинга в продвижении услуг и товаров, а, следовательно, создание новых рабочих мест из-за повышения востребованности специалистов данной отрасли [2].

Социальные сети – одновременно и огромная база данных и мощный инструмент эффективной добычи, обработки информации о пользователях, с возможностью детального структурирования по множеству критериев, к примеру, увлечениям, взглядам, интересам и пристрастиям. Эта информации может использоваться для продвижения товаров и услуг, создания рынков сбыта, рабочих места и общественные мнения [1,5].

Привлекая уйму всевозможных членов общества, соц. сети преобразовались из средств коммуникации в гибкие платформы для маркетинга, продажи товаров и реализации услуг, а также всевозможного досуга [5]. Высокая степень интеграции различных платформ и бурное расширение пользовательских баз сопряжено с рисками, к примеру:

- Кража персональных, конфиденциальных или личных данных;
- Создание и распространение ложных новостей или информационных вбросы;
- Вмешательства в голосования, повышения рейтинга тех или иных ресурсов;
- Создание проблем для социального маркетинга и исследований;
- Ведение нелегального бизнеса, хищения денежных средств;
- Вредительство и деструктивные действия на сетевых коммуникациях.

Часто для этих целей применяют – виртуальных пользователей или ботов [4].

Бот – это специальное программное обеспечение, запрограммированное подражать поведению реальных пользователей в социальных сетях [7]. Ботов используют для автоматизации рутинных процессов, массовых вычислений, различных системы поддержки первой линии или имитация реального пользователя, что может использоваться злоумышленниками для получения или распространения информации, продвижения идей либо брендов, создания спроса на определённые категории товаров или услуг, привлечения интереса людей к обсуждению или наоборот отвлечению от какой-либо темы [3].

По алгоритмам своего функционирования виртуальные пользователи делятся на два типа, автоматические и управляемые. Автоматические исполняют простые, заранее прописанные, инструкции (рассылка сообщений, различные накрутки, вступления в группы и тд.). Подобные бот-программы зачастую используются для завышения всевозможных статистических показателей определённых пользователей или групп, а также спам-рассылки сообщений. Управляемые виртуальные пользователи работают под надзором операторов. При регулярном мониторинге и корректировке действий со стороны операторов, реже подвергаются блокировке, так как их уровень социализации выше, они соблюдают установленные в соц. сети правила. Классическим представителями управляемых ботов являются страницы интернет магазинов или фирм [2,5].

По назначению боты возможно разделит несколько типов:

- Игровые - используются для накрутки рейтинга приложений в социальных сетях и имитации активности в них;
- Новостные - занимаются размещением последних новостей на своей странице. Относительно безвредные, но могут быть использованы для вброса или дезинформации;
- Маскирующиеся под реальных людей, распространяя сообщения реальный пользователей. Используются аналогично с новостными;
- Аккаунты, публикующие чужие новости и репосты, используются для имитации активности и создания всевозможных накруток и повышения спроса [7].

Виртуальные пользователи, боты, в настоящее временя, создают уйму проблем, как для собственников социальных сетей, из-за активности виртуальных пользователей растёт нагрузка на инфраструктуру и коммуникации сетей, что может привести к проблемам доступа к сетям, для рядовых пользователей, постоянные спам-сообщения и дезинформация, так и для представителей бизнеса или исследовательской деятельности, затрудняют исследования рынка или социальных течений на основе реальных нужд, интересов и потребностей социума [1].

Последствиями активности ботов являются:

- дезинформация общественности;
- необъективность исследований маркетинга (SMM);
- утечки персональных и конфиденциальных данных пользователей.

Всё вышеперечисленное в свою очередь приводит к дискредитации социальных сетей с последующим оттоком активных пользователей[5].

Противодействие виртуальным пользователям в социальных сетях в текущий момент времени осложнено высокой мобильностью и оперативностью взломщиков, а также политикой социальных сетей по общедоступности среднестатистическому пользователю. Реализуются в основном минимальные превентивные действия препятствующие регистрации новых ботов:

- CAPTCHA и обратные тесты Тьюринга - используется для определения является пользователь бот-программой или человеком и затрудняющие использующие технологий распознавания графической информации в текстовую или требующие смыслового анализа изображений или фраз [7];
- Проверка пользователя на подлинность через отправку на телефонный номер сообщения с кодом подтверждения, который затем вводится в определённое поле при аутентификации или регистрации пользователя;
- Использование специализированных программ для аутентификации (к примеру, Google Authenticator), процедурно генерирующих дополнительный, помимо стандартных логина и пароля, ключ, доступный в краткий промежуток времени (до полуминуты);
- Ограничение пропускной способности или лимитированное количества запросов в системе за определённый промежуток времени.

При распознавании ботов, из-за особенностей человеческой личности, существует проблема стопроцентного определения является ли пользователь реальным, а не виртуальным. Причина этого не столько в активном развитии алгоритмов работы ботов, а скорее в инертности и малой, а порой и минимальной, активности пользователей, использующих соц. сеть для каких-либо утилитарных функциях, к примеру, сугубо прослушивание музыки. Подобных пользователей не абсолютное большинство, но число их велико чтобы просто проигнорировать как погрешность [3]. Имея подобные исходные данные необходимо применять понятия нечёткой логики и для более точного определения бот-программ стоит опираться не одну или пару характеристик, а на совокупность многих, зачастую не связанных, признаков, выявленных заблаговременно. Руководствуясь принципами кластеризации и классификации лучшим вариантом будет использование рейтинговой системы: профиль пользователя будет признан ботом только согласно суммарной оценке критериев, заданных с определёнными балами за различные категории, после анализа профиля [6].

Таблица 1. Признаки виртуальных пользователей для анализа профилей пользователей

Заполнение профиля	При регистрации бота вручную заполняют минимум данных, экономя время, при программной регистрации данные заполняются максимально, согласно форме.
Список друзей.	Среднестатистические пользователи имеют от 20 до 200 «друзей». Для массовости и охвата аудитории боты добавляют в друзья всех подряд. При наличии у профиля более 300 «друзей», есть вероятность, что это бот. Но существуют реальные пользователи, у которых отсутствуют друзья и подписчики, но они активны, ведут записи, комментируют чужие посты, выкладывают фотографии. В данном аспекте важно не число друзей, а их «качество» или отношение времени существования профиля к числу публикаций и количество входящих и исходящих комментариев.
Список подписчиков.	Обычно профили пользователей имеют от 5 до 50 подписчиков, туда попадают все профили либо не добавляемые, либо удаляемые пользователем, но в любом случае их количество не должно превышать 300-500 профилей.
Блокировка профиля.	Заблокированные профили в социальной сети, как правило, не важна причина блокировки, ведь реальный пользователь восстановит свой аккаунт, а владельцу бота проще создать несколько новых ботов.
Верифицированные профили	Как правило, такие аккаунты получают публичные личности, лично подтверждаемые.
Главная фотография профиля.	Не должно содержаться изображение со сторонней рекламой, так же не должна являться единственной, иначе с большой вероятностью это бот.
Агрегация с соц. сетями и сервисами.	Отсутствие ссылок на агрегацию с другими соц. сетями или сервисами пользователя может являться признаком бота.
Публикации на странице.	Часто бота маскируют, наполняя профиль некоторым количеством публикаций. При подобной ситуации следует учитывать отношение количества личных публикаций к общему количеству записей и промежутки между активностями, а также отношение времени существования профиля к количеству публикаций. Отсутствие собственных публикаций в профиле может быть признаком бота.
Участие в сообществах.	Участие профиля в большом количестве групп разных городов на всевозможные тематики будет признаком виртуального пользователя, так как реальным людям

	сложно одновременно следить и поддерживать интерес к новостям множества не связанных сообществ, а вот боту выгодно находиться в как можно большем их количестве для охвата аудитории.
Наличие орфографических ошибок.	Совершение ошибок свойственно всем людям. Если в профиле с момента его создание не было ни единой ошибки – это косвенно может быть признаком бот-программы.
Активность других пользователей на странице профиля.	У реальных пользователей всегда есть, но может быть не большая, активность «друзей», а также количество просмотров записей, в среднем не больше списка друзей. Если активность, в основном, исходит от постороннего пользователя возможно это бот.
Обилие рекламных записей или ссылок.	Редко реальные пользователи размещают на своей странице много рекламных записей или специально оставляют вредоносные ссылки, так как это может привести к блокировке аккаунта.
Содержание сообщений и комментариев.	Использование в комментариях и сообщениях рекламы, неизвестных ссылок, примитивов или не соответствие теме, являются могут признаками виртуального пользователя. Наибольшее распространение, даже у прогрессивных ботов, имеют формы «вопрос-ответ».
Скорость написания комментариев.	Реальные пользователи не могут комментировать записи быстрее, чем один комментарий в секунду, ведь необходимо прочитать, что комментируешь, связно придумать ответ и используя устройства ввода записать его. Нужно минимум 10 секунд.

Все выше перечисленные характеристики стоит рассматривать только в комплексе. В текущей момент времени разработка методов анализа профилей и распознание среди них виртуальных пользователей весьма актуальна, особенно в рамках защиты персональных и конфиденциальных данных согласно реализуемому федеральному закону Российской Федерации от 27 июля 2006 года № 152-ФЗ «О персональных данных». Программы по детектированию виртуальных пользователей социальных сетей позволит повысить объективность маркетингового анализа и социальных исследований, голосований, проводимых в соц. сетях, существенно сократить массовые утечки персональных данных, общее количество ложных новостей и дезинформации, создаст дополнительные препоны для ведения нелегального бизнеса, а также повысит защищённость сетевых коммуникаций.

Библиографический список

1. Carrington P., Scott J., Wasserman S. Models and Methods in Social Network Analysis. Cambridge: Cambridge University Press, 2005.
2. Global social media research summary 2019 // smartinsights.com URL: https://www.smartinsights.com/social-media-marketing/ (дата обращения: 13.03.2019).
3. Pallis G., Zeinalipour-Yazti D., Dikaiakos M. D. Online Social Networks: Status and Trends. New Directions in Web Data Management 1 // Studies in Computational Intelligence. 2011. Т. 331. С. 213-234.
4. База уязвимостей интернет ресурсов // en.securitylab.ru URL: https://en.securitylab.ru/nvd/ (дата обращения: 12.03.2019).
5. Давыдов А.А. Системная социология: Social Networks Mining // isras.ru URL: http://www.isras.ru/?page_id=1033 (дата обращения: 11.03.2019).
6. Кукушкин И.М., Лясин Д.Н. Разработка системы типологизации пользователей web-ресурсов // Постулат. 2018. №6 (32). С. 36.
7. Свободная энциклопедия // wikipedia.org URL: https://ru.wikipedia.org/wiki/ (дата обращения: 12.03.2019).