Научная статья
УДК 004:519.854
DOI:10.31854/1813-324X-2023-9-1-94-104

Метрики вредоносных социальных ботов

Максим Вадимович Коломеец[1], kolomeec@comsec.spb.ru
Андрей Алексеевич Чечулин[1, 2] ✉, chechulin.aa@sut.ru

[1]Санкт-Петербургский Федеральный исследовательский центр Российской академии наук,
Санкт-Петербург, 199178, Российская Федерация
[2]Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича,
Санкт-Петербург, 193232, Российская Федерация

Аннотация: *В работе представлена параметризация вредоносных ботов с помощью метрик, которые могут быть основой для построения моделей распознавания параметров ботов и качественного анализа характеристик атак в социальных сетях. Предложен ряд метрик для описания характеристик ботов социальной сети ВКонтакте, а именно: доверие, выживаемость, цена, тип продавца, скорость и экспертное качество. Для извлечения данных метрик разработан подход, который основан на методиках контрольной закупки и теста Тьюринга. Основное преимущество данного подхода состоит в том, что он предлагает извлекать признаки из данных, полученных экспериментальным способом, и тем самым получить более обоснованную оценку в сравнении с экспертным подходом. Также работа содержит описание эксперимента по извлечению метрик вредоносных ботов социальной сети ВКонтакте с использованием предложенного подхода, и результаты анализа зависимости метрик. Эксперимент подтверждает возможность извлечения и анализа метрик. В целом, предложенные метрики и подход к их извлечению могут стать основой для перехода от бинарного обнаружения атаки в социальных сетях к качественному описанию атакующего и его возможностей, а также анализу эволюции ботов.*

Ключевые слова: *безопасность социальных сетей, социальные боты, социальная инженерия, метрики, дезинформация, фейковые аккаунты, анализ рисков*

Источник финансирования: Исследование выполнено за счет гранта Российского научного фонда № 18-71-10094.

Ссылка для цитирования: 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

Properties of Malicious Social Bots

Maxim Kolomeets[1], kolomeec@comsec.spb.ru
Andrey Chechulin[1, 2] ✉, chechulin.aa@sut.ru

[1]Saint-Petersburg Federal Research Center of the Russian Academy of Sciences,
St. Petersburg, 199178, Russian Federation
[2]The Bonch-Bruevich Saint-Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Abstract: *The paper considers the ability to describe malicious bots using their characteristics, which can be the basis for building models for recognising bot parameters and qualitatively analysing attack characteristics in social networks. The following metrics are proposed using the characteristics of VKontakte social network bots as an example: trust, survivability, price, seller type, speed, and expert quality. To extract these metrics, an approach is*

proposed that is based on the methods of test purchases and the Turing test. The main advantage of this approach is that it proposes to extract features from the data obtained experimentally, thereby obtaining a more reasonable estimation than the expert approach. Also, an experiment on extracting metrics from malicious bots of the VKontakte social network using the proposed approach is described, and an analysis of the metrics' dependence is carried out. The experiment demonstrates the possibility of metrics extracting and analysis. In general, the proposed metrics and the approach to their extraction can become the basis for the transition from binary attack detection in social networks to a qualitative description of the attacker and his capabilities, as well as an analysis of the evolution of bots.

Keywords: *social media security, social bots, social engineering, metrics, disinformation, fake accounts, risk analysis*

Funding: The study was supported by the Russian Science Foundation grant No. 18-71-10094.

For citation: Kolomeets M., Chechulin A. Properties of Malicious Social Bots. *Proc. of Telecom. Universities.* 2023;9(1):94–104. (in Russ.) DOI:10.31854/1813-324X-2023-9-1-94-104

Введение

Вредоносные боты являются одним из основных инструментов проведения атак в социальных сетях. Несмотря на существенный прогресс в развитии методик выявления и противодействия ботам, методы определения их характеристик развиты недостаточно для их практического применения. Данная работа направлена на изучение видов ботов социальных сетей, а также способах их параметризации, что также позволяет описать свойства атаки ботов в социальных сетях и возможностей атакующего. При этом, в основе исследования лежит явление возрастающей гетерогенности ботов – когда боты начинают «эволюционировать» по различным параметрам, что сказывается как на способности их детектирования, так и на возможности причинения ущерба.

Для того, чтобы успешно противодействовать таким атакам, когда используются боты различных видов, необходимо систематизировать функционал и особенности ботов в виде метрик, от которых может зависеть набор выбранных контрмер или выводы расследования уже совершенной атаки. Учитывая особенности функционирования ботов, предлагается подход к извлечению метрик ботов, который построен на двух методиках: на базе контрольной закупки и на основе теста Тьюринга. Научная новизна предложенного подхода лежит в концепции того, что вредоносные боты подлежат качественному описанию в виде метрик, которые в последствии можно использовать в системах противодействия, анализе рисков и расследованиях атак постфактум. Научная значимость состоит в предложенных методиках извлечения метрик, а практическая значимость в самих метриках, которые можно использовать в существующих системах выявления ботов.

Для поддержки результатов и выводов данной концепции, в работе также приводятся алгоритмы расчета метрик, эксперименты по расчету метрик вредоносных ботов социальной сети ВКонтакте.

Данная работа состоит из следующих разделов: второй раздел описывает контекст научной проблемы и релевантные исследования – обсуждается проблема параметризации ботов в контексте анализа и противодействия атакам в социальных сетях; в третьем разделе представлена методика на основе контрольной закупки, которая позволяет рассчитать такие метрики ботов как *цена*, *скорость*, *тип продавца*, *экспертное качество* и *выживаемость* бота; в четвертом разделе представлена методика на основе теста Тьюринга, которая позволяет рассчитать метрику *доверия* и ее вариации; в пятом разделе представлена экспериментальная часть по расчетам метрик ботов социальной сети ВКонтакте; в шестом разделе приводится анализ рассчитанных метрик и обсуждается их интерпретация; последний раздел – это заключение.

Контекст научной проблемы и релевантные исследования

В целом существует множество определений ботов, которые включают в себя social bots, spambots, sybils, autobots, trolls, и др. Данная несогласованность в терминологии связана с тем, что ранее под ботами понимали такие аккаунты социальных сетей, которые действуют исключительно автоматически и у которых характер взаимодействий с легитимными пользователями ограничен определенным алгоритмом, а прочие виды вредоносных аккаунтов подразделяли в отдельные виды. Постепенное усложнение [1–3] поведения ботов создало большое количество их разновидностей, которые выходят за пределы первоначального определения. В данном исследовании боты определяются как аккаунты социальных медиа, которые используют для социоинженерных атак [4]. Результатом таких атак является создание процесса, явления и отдельного события, которые не могут образоваться естественным образом. Такие аккаунты действуют не по воле человека, которого данный аккаунт представляет.

Таким образом, в данную категорию попадают:

1) взломанные аккаунты – когда злоумышленник получает доступ к чужому аккаунту и от имени предыдущего владельца осуществляет искусственные воздействия в социальной сети;

2) специально созданные аккаунты – когда злоумышленник генерирует или крадет контент, и на его основе создает фейковую идентичность, от имени которой осуществляет искусственные воздействия в социальной сети;

3) аккаунты, владельцы которых подкуплены злоумышленником – когда пользователи получают деньги или другие блага за осуществление искусственных воздействий в социальной сети.

При этом, необходимо учитывать, что не все боты вредоносны. Например, под вышеперечисленные критерии не попадают аккаунты ботов коммерческих компаний, выдуманных персонажей поп-культуры и ботов-сервисов. Данные аккаунты не несут угроз пользователям.

Ключевыми признаками [4] вредоносности ботов является:

– сокрытие факта того, что аккаунт является ботом (вредоносные боты стремятся выдавать себя за реально существующих людей, так как доверие пользователя является одним из элементов успешности проведения атаки, в то время как невредоносные боты не стремятся обмануть пользователя);

– инициатива по отношению к пользователям (вредоносные боты первыми осуществляют взаимодействие с пользователем, в то время как невредоносные боты лишь отвечают на запросы пользователя;

– искажение естественных процессов (вредоносные боты стремятся исказить естественные либо имитировать социальные процессы, в то время как невредоносные боты предоставляют сервисы).

При этом многие исследователи отмечают явление эволюции ботов [2, 5] – когда с ростом научно-технических возможностей атакующего, возрастает сложность поведения ботов [6–7] и методов их создания [8], что в свою очередь влияет на возможности детектирования и противодействия [3, 5, 9]. Наиболее наглядно эволюция ботов проявляется в дифференциации серого рынка ботов [8, 10] – когда покупатель может выбрать ботов определенного вида из имеющихся на рынке предложений, чтобы придать атаке желаемые свойства (например, сопротивляемость ботов блокировке, большая схожесть с целевой аудиторией и др.).

Другая проблема, будучи производной особенности эволюции ботов, состоит в том, что наиболее популярным способом выявления ботов является использование методов машинного обучения с учителем [5]. Несмотря на развитие и других методов, например, машинного обучения без учителя [11–12] или аналитической статистики [13–15], данные методы уступают в результативности [5]. При этом одной из ключевых проблем существующих методов обучения с учителем является отсутствие корректных наборах данных, на основе которых обучаются и/или тестируются разрабатываемые модели.

На основе 41 публикации [5] на тему обнаружения ботов, были выделены следующие методы разметки, которые используют исследователи.

1) Ручная разметка. Эксперт осуществляет разметку на основе своего мнения, таким образом, ее качество напрямую зависит от способности эксперта распознать бота (*англ*. True Positive Rate) и не принять реального пользователя за бота [16] (*англ.* True Negative Rate).

2) Закупка. Боты закупаются у продавца ботов, таким образом, исследователь может быть уверен в качестве разметки.

3) Аномальное поведение. Разметка осуществляется на основе формальных признаков аномального поведения аккаунтов. Качество разметки зависит от видов ботов, и от того, насколько сложное поведение они могут воспроизводить.

4) Декларация. Разметка осуществляется третьей стороной на основе другого средства обнаружения ботов. Например, исследователи могут отнести к ботам все заблокированные социальной сетью аккаунты, в таком случае разметка осуществляется функционирующей системой защиты социальной сети. Таким образом, качество разметки напрямую зависит от эффективности средства обнаружения ботов, а разрабатываемые на основе таких данных решения не смогут превысить точность используемого метода обнаружения.

Также некоторые исследователи не указывали, какой метод они использовали, или использовали свой собственный метод разметки. Также получить наборы ботов можно на основе инсайдерской информации – например, при получении доступа к отчетам продавцов ботов. Популярность того или иного метода разметки по проанализированным работам представлена на рисунке 1.

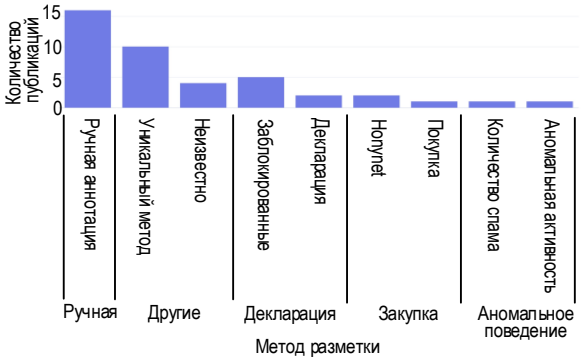

Рис. 1. Популярность методов разметки ботов

Fig. 1. The Popularity of Bot Markup Methods

Единственным надежным методом, который позволяет получить достоверную разметку (*англ.* Ground Truth Data), является метод закупки и получение меток на основе инсайдерской информации. Среди проанализированных работ метод закупки использовался один раз. Применение других методов приводит к тому, что в наборе оказываются легитимные пользователи, а некоторые виды ботов могут отсутствовать, и таким образом, модель не сможет их распознать.

Одним из путей решения вышеперечисленных проблем является переход от бинарного представления (бот/не бот) к описанию ботов, основанном на объективно извлекаемых метриках, способных описать свойства ботов. При использовании метрик становится возможным: во-первых, качественно описывать процесс эволюции ботов – как меняются метрики ботов, найденных в результате мониторинга; во-вторых, качественно описывать свойства атак – какие характеристики имеют боты, участвующих в атаке, и как это влияет на ущерб и защищенность; в-третьих, дифференцировать оценку эффективности методов обнаружения – какие виды ботов система обнаружения распознает лучше и какие – хуже.

Методика извлечения характеристик ботов на основе метода закупки

Данная методика основана на создании фейкового сообщества, для которого в контролируемых исследователем условиях закупаются боты и исключаются реальные пользователи. Таким образом становится возможным получить чистые наборы данных, гарантированно состоящие исключительно из ботов – Ground Truth Labels, а также определить метрики ботов, исходя из процесса сбора наборов данных. Данная методика состоит из 6 шагов, а сами метрики выделены полужирным курсивом.

Шаг 1. Создание списка продавцов ботов и их услуг с указанием описания набора ботов от продавца [8]. Как правило, один продавец предлагает на выбор ботов нескольких разных качеств. Список включает:

– название продавца;

– описание ботов от продавца;

– качество ботов, декларируемое продавцом (может крайне отличаться, например, «Отличное», «Ультима», «Живые», и др.);

– ***экспертное качество*** ботов по категориальной шкале [НИЗКОЕ, СРЕДНЕЕ, ВЫСОКОЕ], определяемой экспертом на основании описания продавца и качества декларируемым продавцом;

– ***тип продавца*** по шкале по категориальной шкале [МАГАЗИН, БИРЖА]; магазины – продавцы ботов, у которых можно купить вредоносную активность, исходя из предложения, биржи ботов – площадки, где размещается запрос о вредоносной активности и любой продавец (или частный владелец аккаунта) может выполнить запрос или его часть за вознаграждение;

– вид активности (лайк, комментарий и др. – зависит от вида социальной сети);

– ***цена*** за единицу активности бота (количественная шкала [0, +INF] в рублях на день покупки).

Шаг 2. Создание фейкового сообщества и наполнения его контентом. Сообщество должно удовлетворять двум условиям:

– должно выглядеть настоящим, чтобы не вызвать у продавца ботов подозрений; для этого сообщество наполняется контентом (фотографии, посты, другие фейковые пользователи); некоторые продавцы ботов отказывают в услугах, если в сообществе состоит мало людей, либо у сообщества низкая активность;

– сообщество не должно быть привлекательным для реального пользователя, чтобы исключить вероятность того, что реальные пользователи проявят активность в сообществе; для этого общество должно обладать абсурдной / непривлекательной тематикой (например, перевозки между несуществующими городами и др.).

Шаг 3. Закупка в сообщество партии ботов (один продавец, одно качество) и сохранение идентификаторов аккаунтов, осуществивших активность с сохранением метки продавца и качества для данной партии. При закупке продавцу дается задание, например – поставить *N* лайков посту *X*. По завершении закупки также размечается ***скорость*** ботов, как разница во времени между моментом оплаты и завершением задачи. Скорость размечается как категориальная мера по шкале [МОМЕНТАЛЬНО, ЧАС, СУТКИ] – задание завершено:

– моментально (менее минуты);

– за несколько часов;

– за сутки и более;

Шаг 4. Удаление активности в сообществе.

Шаг 5. Повторение шага 3 для других продавцов ботов и предоставляемых ими видов аккаунтов.

Шаг 6. Ожидание некоторого времени (несколько недель или месяцев), и последующая проверка того, какое соотношение ботов из определенного набора оказалось заблокированным системами защиты социальной сети. Выражает ***выживаемость*** (количественная шкала [0, 1] как вероятность блокировки).

В результате формируется множество наборов данных, где каждый элемент множества представляет собой одно предложение от продавца ботов. Каждый набор данных, в свою очередь, имеет следующие идентифицирующие его свойства:

– название продавца;

– качество ботов, декларируемое продавцом.

Кроме общих свойств, каждый набор данных содержит список аккаунтов с метриками:

– id аккаунта бота;

– заблокирован ли аккаунт [ДА, НЕТ];

– вид активности: [лайк, комментарий и др.];

– ***экспертное качество бота***, по категориальной шкале [НИЗКОЕ, СРЕДНЕЕ, ВЫСОКОЕ];

– ***тип продавца***, по категориальной шкале [МАГАЗИН, БИРЖА];

– ***цена*** за единицу активности, количественная шкала [0, +INF];

– ***скорость***, категориальная шкала [МОМЕНТАЛЬНО, ЧАС, СУТКИ];

– ***выживаемость,*** количественная шкала [0, 1].

Методика извлечения характеристик ботов на основе теста Тьюринга

Одним из свойств бота является стремление маскировать свои действия под активность реальных пользователей. От этого зависит успех атаки, так как пользователь не станет взаимодействовать с аккаунтом, если будет знать, что это бот. Например, он не будет верить отзыву о товаре, не будет передавать банковские данные, или не будет вступать в дискуссию, если известно, что собеседник – бот. Методика предлагает рассчитать метрику доверия, которая будет описывать способность человека распознавать анализируемый вид ботов. Для того, чтобы получить эту оценку, была создана методика на основе теста Тьюринга [17].

Методика основывается на разметке наборов ботов аннотаторами. За счет разницы между ответами аннотаторов и реальными лейблами можно определить способность человека распознать ботов анализируемого вида – по аналогии с тестом Тьюринга. Для того чтобы оценить эту разницу, был разработан следующий дизайн теста:

1) наборы для разметки формируются из трех наборов по X случайных пользователей:

– репрезентативных (случайные пользователи генеральной совокупности всех пользователей социальной сети);

– смещенных (пользователи, проявившие какую-либо активность: в сообществе, чате и др.);

– верифицированных (аккаунты, проверенные исследователем);

2) из Y наборов по Z случайных ботов.

Таким образом, один набор будет содержать $3 * X + Y * Z = N$ аккаунтов, среди которых необходимо обнаружить ботов. Использование трех разных наборов пользователей связано с тем, что свойства пользователей в наборе влияют на ответы аннотаторов. Например, распознать ботов среди верифицированных пользователей проще, в сравнении со случайными аккаунтами. Данные 3 набора гарантируют, что в итоговый набор попадут пользователи различной схожести с ботами и с разной степенью социальной гомофилии [18], что соответствует основным сценариям анализа (вся социальная сеть, сообщество, близкая группа людей). Y обеспечивает ситуацию, когда всякий аннотатор разметит хотя бы по одного боту из каждого набора.

Таким образом, во-первых, итоговый набор из N аккаунтов перемешивается и подается на разметку одному аннотатору. Во-вторых, каждый ответ аннотатора сохраняется в файл с ответами с идентификатором размеченного аккаунта и одним из ответов аннотатора:

– бот (данный аккаунт является ботом);

– пользователь (аккаунт является реальным пользователем);

– я не знаю (аннотатор сомневается в ответе).

Для каждого набора можно рассчитать метрики эффективности обнаружения аннотаторами ботов определенного набора (таблица 1).

ТАБЛИЦА 1. Метрики эффективности обнаружения ботов пользователями

TABLE 1. Performance Metrics of Bot Recognition by Users

Ответ аннотатора	Тип аккаунта	Метрика
Бот распознан корректно (успех)	все аккаунты	TP
	аккаунт заблокирован	TP_Z
	аккаунт не заблокирован	TP_{NZ}
Бот распознан некорректно (провал)	все аккаунты	FN
	аккаунт заблокирован	FN_Z
	аккаунт не заблокирован	FN_{NZ}
Ответ «я не знаю»	все аккаунты	IDN
	аккаунт заблокирован	IFN_Z
	аккаунт не заблокирован	IDN_{NZ}

На основе метрик таблицы 1 можно рассчитать метрики доверия, которые описывают способность пользователя распознать бота (таблица 2).

ТАБЛИЦА 2. Метрики доверия и способы их расчета

TABLE 2. Trust Metrics and How to Calculate Them

Метрика	Формула	Обработка	
		ответов «не знаю»	заблокированных аккаунтов
$Trust$	$\frac{TP}{TP+FN}$	не учитываются	учитываются
$Trust^{IDN}$	$\frac{TP}{TP+FN+IDN}$	учитываются как провал	учитываются
$Trust_{NZ}$	$\frac{TP_{NZ}}{TP_{NZ}+FN_{NZ}}$	не учитываются	не учитываются
$Trust_{NZ}^{IDN}$	$\frac{TP_{NZ}}{TP_{NZ}+FN_{NZ}+IDN_{NZ}}$	учитываются как провал	не учитываются
$Trust_Z$	$\frac{TP+FN_Z}{TP+FN}$	не учитываются	учитываются как успех
$Trust_Z^{IDN}$	$\frac{TP+FN_Z+IDN_Z}{TP+FN+IDN}$	учитываются как провал	учитываются как успех

Примечание. В таблице: $Trust$ – доверие (здесь и далее); $Trust^{IDN}$ – учитывает ответы «не знаю»; $Trust_{NZ}$ – учитывает только незаблокированные аккаунты; $Trust_{NZ}^{IDN}$ – учитывает ответы «не знаю» и только незаблокированные аккаунты; $Trust_Z$ – незаблокированные аккаунты считаются ботами независимо от ответа; $Trust_Z^{IDN}$ – учитывает ответы «не знаю», а незаблокированные аккаунты считаются ботами независимо от ответа.

Для оценки доверия были собраны 100 000 случайных аккаунтов, которые считались как репрезентативный список пользователей. Для формирования набора смещенных пользователи были взяты аккаунты из набора ботов MKVK2021 [19], в которых содержатся пользователи, проявившие активность в одной из групп ВКонтакте. В качестве верифицированных пользователей использовались аккаунты студентов из исследования [20].

Экспериментальная проверка методик

С использованием методики контрольной закупки были собраны 65 размеченных наборов ботов от 25 компаний, предлагающих услуги активности ботов в социальной сети ВКонтакте. Экспертное качество было размечено тремя аннотаторами, исходя из описаний продавцов. Время ожидания для оценки выживаемости составило 3 месяца. В совокупности были собрано 22325 аккаунтов ботов, из которых 18444 аккаунтов являются уникальными. Итоговое распределение метрик изображено на рисунке 2.

Для разметки аккаунтов аннотаторами был реализован бот в мессенджере Telegram, где можно в удобной форме размечать аккаунты. При запуске бота, для каждого аннотатора индивидуально формировался набор из 101 аккаунта (по 12 – из репрезентативного, смещенного и верифицированного набора пользователей и по одного боту из 65 наборов) и поочередно предлагалось оставить метку для аккаунта. При завершении разметки 101 аккаунта, формировался еще один набор; таким образом аннотатор смог разместить больше 101 аккаунта. В качестве аннотаторов выступили 30 студентов СПбГУТ им. проф. М.А. Бонч-Бруевича. В результате эксперимента были получены 3168 меток. Метрики по каждому аннотатору представлены на рисунке 3, где ось *X* представляет скрытые имена аннотаторов.

Для расчета метрики доверия таким же образом были рассчитаны метрики эффективности обнаружения для каждого набора, которые схематично объяснены на рисунке 4.

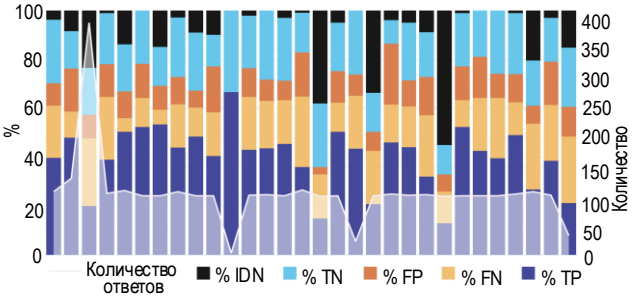

Рис. 3. Метрики эффективности обнаружения по аннотаторам

Fig. 3. Metrics of Detection Efficiency by Annotators

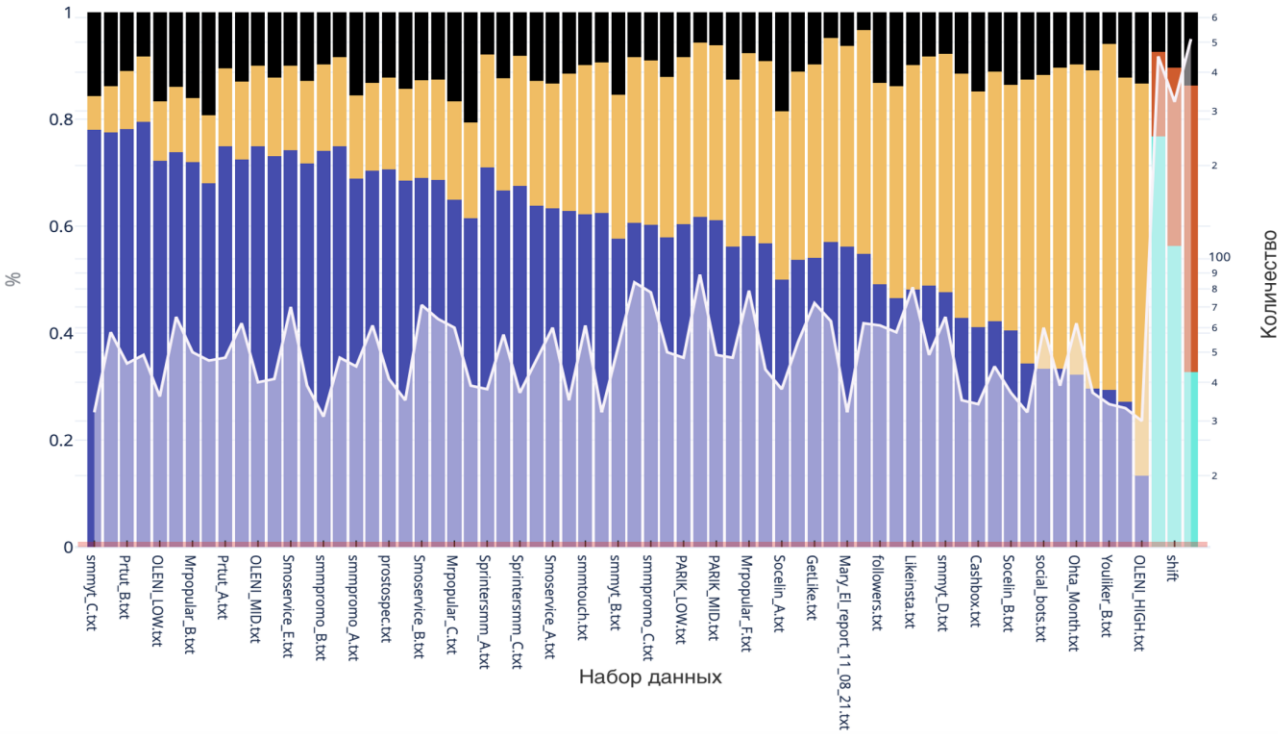

Рис. 4. Метрики эффективности обнаружения для наборов данных

Fig. 4. Metrics of Detection Efficiency for Datasets

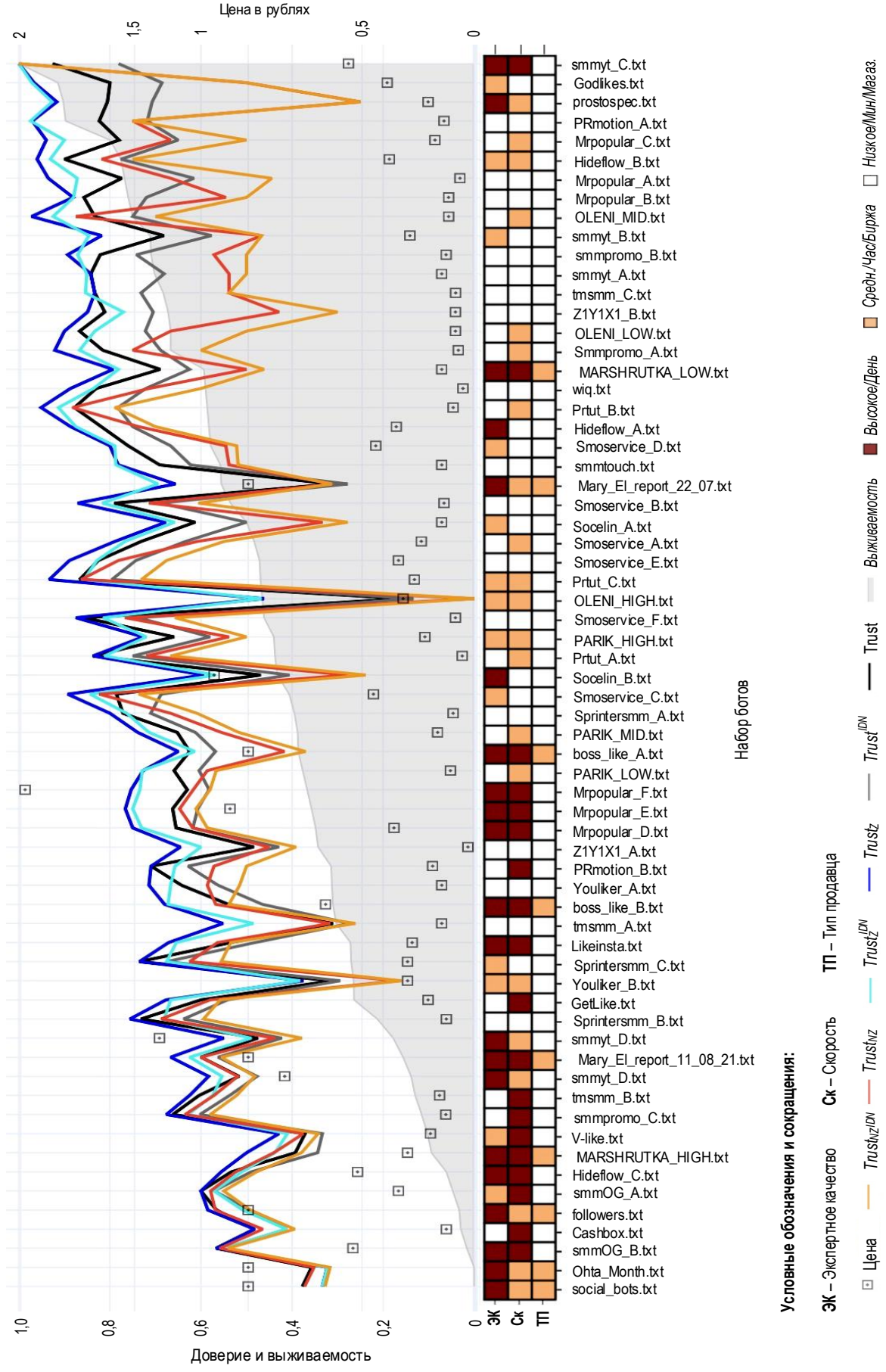

Рис. 2. Распределение метрик ботов социальной сети ВКонтакте

Fig. 2. Distribution of Bot Metrics for VKontakte Social Network

Как видно из рисунка 4, для наборов, представленных в левой части, пользователи смогли выявить около 80 % всех ботов, в то время как в правой части – ниже 30 %. Исходя из данных метрик, и были рассчитаны метрики доверия, представленные на рисунке 2.

Анализ подхода

Для того, чтобы иметь возможность характеризовать атаки, в социальных сетях был разработан ряд метрик ботов, которые можно использовать как характеристики атаки. Несмотря на их большое разнообразие, они были построены только на тех характеристиках, которые возможно однозначно извлечь из наборов данных: данные метрики рассчитываются на основе закупок ботов, экспериментов, экспертной разметки, и аннотации, произведенной в ходе сбора наборов ботов.

Метрики из таблицы 3 можно интуитивно интерпретировать следующим образом.

Доверие. Выражает качество бота с точки зрения способности человека распознать бота. Представляет собой ряд метрик, рассчитанных на основе экспериментов по выявлению ботов людьми – ответов аннотатора с учетом заблокированных аккаунтов и ответов «я не знаю». Из рисунка 2 следует, что вариации метрики доверия сильно коррелируют между собой, поэтому на практике целесообразно использовать только некоторые из них. Данная метрика наиболее хорошо характеризует вероятность успеха атаки, так как ботов с низким значением метрики жертве сложнее распознать.

Выживаемость. Доля незаблокированных аккаунтов в наборе. Выражает качество бота с точки зрения его способности противостоять блокировке средствами защиты социальной сети. Рассчитывается на основе анализа аккаунта и представляет собой процентное отношение. Данная метрика наиболее хорошо характеризует способность бота противостоять используемым на момент анализа методам обнаружения.

Цена. Стоимость бота в рублях. Выражает стоимость атаки. Рассчитывается на основе стоимости, уплаченной в результате закупки ботов. Данная метрика наиболее хорошо характеризует возможности атакующего.

Тип продавца. Выражает стратегию управления ботами [4, 8] и включает 2 значения: магазин и биржа. Боты из магазинов чаще всего управляются автоматически и создаются в автоматическом или полуавтоматическом режиме, а также не всегда могут генерировать сложный текстовый контент. Боты из биржи главным образом управляются людьми и создаются в ручном режиме, и как следствие, могут генерировать сложный текстовый контент. Данная метрика наиболее хорошо коррелирует с подходом к управлению ботами.

Скорость. Выражает скорость атаки, которая зависит от сложности поведения ботов. Автоматические боты могут совершить атаку моментально (в течение нескольких секунд), а боты, имитирующих естественное поведение, совершают атаку в течение часа или суток. Вручную управляемые боты могут совершать атаку более суток. Данная метрика наиболее хорошо характеризует маневренность атакующего и характер управления ботами.

Экспертное качество. Отражает мнение эксперта, интерпретирующего рекламное описание ботов: насколько сложным в контексте создания и управления является бот с точки зрения продавца. Данная метрика наиболее хорошо характеризует сложность технологий, используемых для создания бота или управления им.

Исходя из их нормирования экспериментальных данных по количественной шкале, что представлено в таблице 3, были также рассчитаны корреляции Спирмена между метриками ботов.

ТАБЛИЦА 3. Пример нормирования метрик ботов

TABLE 3. Bot Metrics Normalization Example

Метрика	Шкала	Интерпретация	
		Чем меньше значение	Чем больше значение
Доверие (*Trust*)	[0:1]	Пользователи доверяют больше	Пользователи доверяют меньше
Выживаемость	[0:1]	Вероятность блокировки ниже	Вероятность блокировки выше
Цена	[0: inf]	Дешевая атака	Дорогая атака
Тип продавца	МАГАЗИН = 0 БИРЖА = 1	Боты из магазина / вероятно, управляются алгоритмически	Боты из биржи / вероятно, управляются вручную
Скорость	МОМЕНТАЛЬНО = 0 ЧАС = 0.5 СУТКИ = 1	Быстрая атака	Медленная атака
Экспертное качество	НИЗКОЕ = 0 СРЕДНЕЕ = 0.5 ВЫСОКОЕ = 1	Боты низкого качества / простые технологии	Боты высокого качества / сложные технологии

Результаты корреляционного анализа для социальной сети ВКонтакте по состоянию на 2022 г. представлены в виде матрицы на рисунке 5, где блоками бирюзового цвета выделены наиболее значимые блоки.

Блок 1. Корреляция экспертного качества между метриками трех экспертов (e1, e2, e3). По данному блоку можно сделать вывод, что мнения экспертов плохо согласуются, и метрика все еще остается сильно субъективной.

Блок 2. Метрика доверия: чем доверие выше (больше вероятность распознать бота) – тем меньше цена, меньше экспертное качество, а сам бот,

вероятно, покупался через магазин, а не биржу. Это в целом сходится с ожидаемыми характеристиками ботов низкого качества. У метрик $Trust_{NZ}$ и $Trust_{NZ}^{IDN}$, из-за того, что при расчете отсутствуют заблокированные аккаунты, пропала корреляция со скоростью бота, что соответствует ожиданиям: автоматические боты, способные совершать быстрые атаки, чаще блокируются соцсетью.

Блок 3. Метрика выживаемости: наборы, в которых много заблокированных аккаунтов, по большей части приобретены в магазинах, а не на биржах, а боты в них обладают высокой скоростью и хорошо распознаются людьми.

Блок 4. Метрики цены, скорости и типа продавца взаимно коррелируют, за исключением корреляции цены и типа продавца.

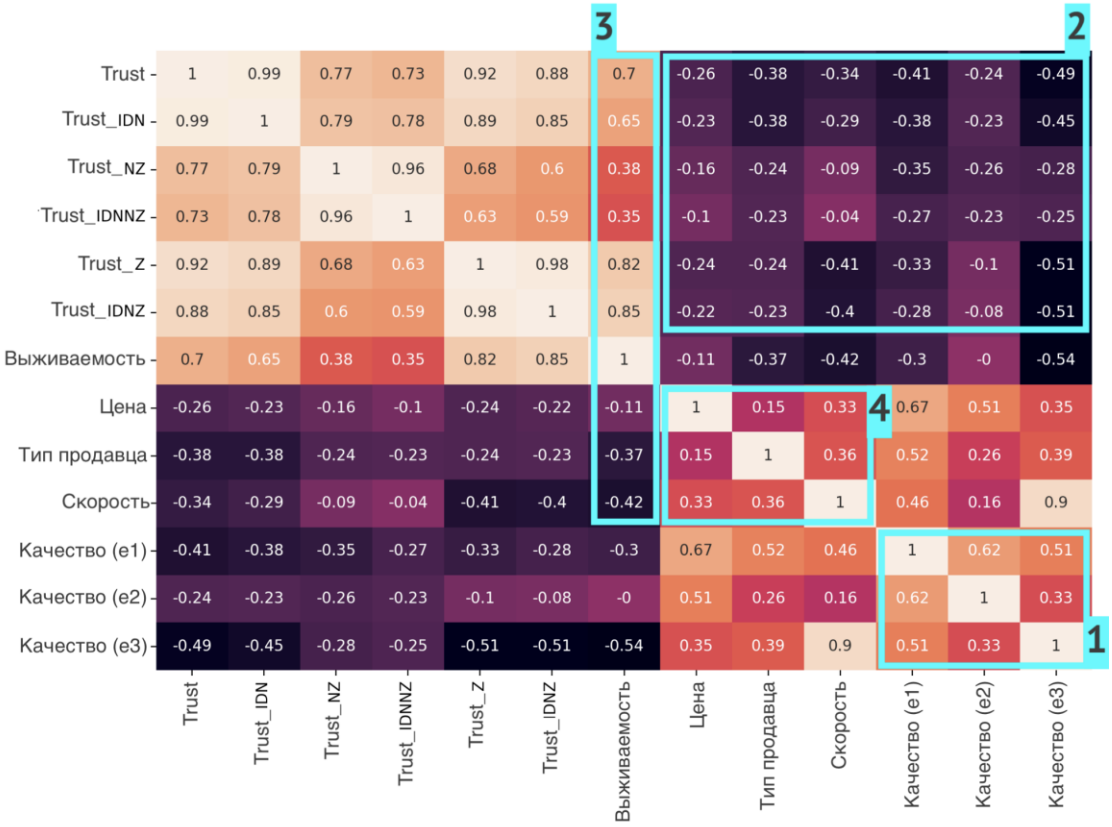

Рис. 5. Матрица корреляций метрик ботов

Fig. 5. Bot Metrics Correlation Matrix

Основной метрикой, влияющей на успешность атаки, является доверие. При этом в ходе корреляционного анализа видно, что метрика не имеет среднюю или сильную корреляцию по Чеддоку с какой-либо другой, полученной в ходе закупки ботов – ценой, скоростью и типом продавца. Сильная и средняя корреляция наблюдается между доверием и выживаемостью – это объясняется тем, что легко распознаваемые людьми боты также легко блокируются применяемыми в социальной сети ВКонтакте системами защиты. В целом можно сделать вывод, что для полноценного описания атаки нельзя обойтись какой-либо одной метрикой.

Заключение

В работе рассмотрена возможность описания вредоносных ботов с помощью следующих метрик: *цена*, *скорость*, *тип продавца*, *экспертное качество*, *выживаемость* и *доверие*. Для их извлечения предложен подход, основанный на методиках контрольной закупки и теста Тьюринга. Эксперименты показали, что с их использованием можно извлечь метрики ботов, позволяющее дифференцировать вредоносные аккаунты на различные виды. Ожидается, что предложенные метрики могут стать, во-первых, основой для качественного описания процесса эволюции ботов – по аналогии с проведенным корреляционным анализом, во-вторых, качественно описывать свойства атак – какие характеристики имеют боты участвующих в атаке и как это влияет на ущерб и защищенность, и в третьих, дифференцировать оценку эффективности методов обнаружения – какие виды ботов система обнаружения распознает лучше и какие хуже.

В дальнейшей работе планируется провести дополнительные эксперименты для выявления дополнительных метрик, характеризующих ботов. Кроме того, планируется использование полученных результатов для выявления каналов, формируемых ботами для распространения информации [21].

Список источников

1. Cresci S. A decade of social bot detection // Communications of the ACM. 2020. Vol. 63. Iss. 10. PP. 72–83. DOI:10.1145/ 3409116

2. Ferrara E., Varol O., Davis C., Menczer F., Flammini A. The rise of social bots // Communications of the ACM. 2016. Vol. 59. Iss. 7. PP. 96–104. DOI:10.1145/2818717

3. Yang C., Harkreader R., Gu, G. Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers // IEEE Transactions on Information Forensics and Security. 2013. Vol. 8. Iss. 8. PP. 1280–1293. DOI:10.1109/TIFS.2013.2267732

4. Vitkova L. Kolomeec M., Chechulin A. Taxonomy and Bot Threats in Social Networks // Proceedings of the International Russian Automation Conference (RusAutoCon, Sochi, Russia, 04–10 September 2022). IEEE, 2022. PP. 814–819. DOI:10.1109/ RusAutoCon54946.2022.9896268

5. Orabi M., Mouheb D., Al Aghbari Z., Kamel I. Detection of Bots in Social Media: A Systematic Review // Information Processing & Management. 2020. Vol 57. Iss. 4. P. 102250. DOI:10.1016/j.ipm.2020.102250

6. Varol O., Ferrara E., Davis C., Menczer F., Flammini A. Online Human-Bot Interactions: Detection, Estimation, and Characterization // Proceedings of the 11th International AAAI Conference on Web and Social Media. 2017. Vol. 11. Iss. 1. PP. 280–289. DOI:10.1609/icwsm.v11i1.14871

7. Stieglitz S., Brachten F., Berthel´e D., Schlaus M., Venetopoulou C., Veutgen D. Do Social Bots (Still) Act Different to Humans? – Comparing Metrics of Social Bots with those of Humans // Proceedings of the 9th International Conference on Social Computing and Social Media. Human Behavior (SCSM 2017, Vancouver, Canada, 9–14 July 2017). Lecture Notes in Computer Science. Vol. 10282. Cham: Springer, 2017. PP. 379–395. DOI:10.1007/978-3-319-58559-8_30

8. Kolomeets M., Chechulin A. Analysis of the Malicious Bots Market // Proceedings of the 29th Conference of Open Innovations Association (FRUCT, Tampere, Finland, 12–14 May 2021). IEEE, 2021. PP. 199–205. DOI:10.23919/FRUCT52173. 2021.9435421

9. Perdana R.S., Muliawati T.H., Alexandro R. Bot spammer detection in twitter using tweet similarity and time interval entropy // Jurnal Ilmu Komputer dan Informasi. 2015. Vol. 8. Iss. 1. PP. 19–25. DOI:10.21609/jiki.v8i1.280

10. The Black Market for Social Media Manipulation. Riga: NATO StratCom COE, 2018.

11. Chavoshi N., Hamooni H., Mueen A. DeBot: Twitter Bot Detection via Warped Correlation // Proceedings of the 16th International Conference on Data Mining (ICDM, Barcelona, Spain, 12–15 December 2016). IEEE, 2016. PP. 817–822. DOI:10.1109/ICDM.2016.0096

12. Dorri A., Abadi M., Dadfarnia M. SocialBotHunter: Botnet Detection in Twitter-Like Social Networking Services Using Semi-Supervised Collective Classification // Proceedings of the 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech, Athens, Greece, 12–15 August 2018). IEEE, 2018. PP. 496–503. DOI:10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00097

13. Vitkova L., Kotenko I., Kolomeets M., Tushkanova O., Chechulin A. Hybrid Approach for Bots Detection in Social Networks Based on Topological, Textual and Statistical Features // Proceedings of the Fourth International Scientific Conference Intelligent Information Technologies for Industry (IITI'19, Ostrava – Prague, Czech Republic, 2–7 December 2019). Advances in Intelligent Systems and Computing. Vol. 1156. Cham: Springer, 2020. PP. 412–421. DOI:10.1007/978-3-030-50097-9_42

14. García-Orosa B., Gamallo P., Martín-Rodilla P., Martínez-Castaño R. Hybrid Intelligence Strategies for Identifying, Classifying and Analyzing Political Bots // Social Sciences. 2021. Vol. 10. Iss. 10. P. 357. DOI:10.3390/socsci10100357

15. Yang K.C., Hui P.M., Menczer F. Bot Electioneering Volume: Visualizing Social Bot Activity During Elections // Companion Proceedings of The 2019 World Wide Web Conference (WWW '19, San Francisco, USA, 13–17 May 2019). New York: Association for Computing Machinery, 2019. PP. 214–217. DOI:10.1145/3308560.3316499

16. Adrian R., Kaiser J. The False positive problem of automatic bot detection in social science research // PLoS ONE. 2020. Vol. 15. Iss. 10. P. e0241045. DOI:10.1371/journal.pone.0241045

17. Boneh D., Grotto A.J., McDaniel P., Papernot N. How Relevant is the Turing Test in the Age of Sophisbots? // IEEE Security & Privacy. 2019. Vol. 17. Iss. 6. PP. 64–71. DOI:10.1109/MSEC.2019.2934193

18. Aiello L.M., Barrat A., Schifanella R., Cattuto C., Markines B., Menczer F. Friendship prediction and homophily in social media // ACM Transactions on the Web. 2012. Vol. 6. Iss. 2. PP. 1–33. DOI:10.1145/2180861.2180866

19. Kolomeets M. Security Datasets – MKVK2021. URL: https://github.com/guardeec/datasets#mkvk2021 (дата обращения 28.02.2023)

20. Branitskiy A., Levshun D., Krasilnikova N., Doynikova E., Kotenko I., Tishkov A., Vanchakova N., Chechulin A. Determination of Young Generation's Sensitivity to the Destructive Stimuli based on the Information in Social Networks // Journal of Internet Services and Information Security. 2019. Vol. 9. Iss. 3. PP. 1–20.

21. Проноза А.А., Виткова Л.А., Чечулин А.А., Котенко И.В., Сахаров Д.В. Методика выявления каналов распространения информации в социальных сетях // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2018. Т.14. № 4. С. 362–377. DOI:10.21638/11702/spbu10.2018.409.

References

1. Cresci S. A decade of social bot detection. *Communications of the ACM*. 2020;63(10):72–83. DOI:10.1145/ 3409116

2. Ferrara E., Varol O., Davis C., Menczer F., Flammini A. The rise of social bots. *Communications of the ACM*. 2016;59(7): 96–104. DOI:10.1145/2818717

3. Yang C., Harkreader R., Gu G. Empirical evaluation and new design for fighting evolving twitter spammers. *IEEE Transactions on Information Forensics and Security*. 2013;8(8):1280–1293. DOI:10.1109/TIFS.2013.2267732

4. Vitkova L. Kolomeec M., Chechulin A. Taxonomy and Bot Threats in Social Networks. *Proceedings of the International*

Russian Automation Conference, RusAutoCon, 04–10 September 2022, Sochi, Russia. IEEE; 2022. p.814–819. DOI:10.1109/RusAutoCon54946.2022.9896268

5. Orabi M., Mouheb D., Al Aghbari Z., Kamel I. Detection of bots in social media: a systematic review. *Information Processing & Management*. 2020;57(4):102250. DOI:10.1016/j.ipm.2020.102250

6. Varol O., Ferrara E., Davis C., Menczer F., Flammini A. Online Human-Bot Interactions: Detection, Estimation, and Characterization. *Proceedings of the 11th International AAAI Conference on Web and Social Media*. 2017;11(1):280–289. DOI:10.1609/icwsm.v11i1.14871

7. Stieglitz S., Brachten F., Berthel´e D., Schlaus M., Venetopoulou C., Veutgen D. Do Social Bots (Still) Act Different to Humans? – Comparing Metrics of Social Bots with those of Humans. *Proceedings of the 9th International Conference on Social Computing and Social Media. Human Behavior, SCSM 2017, 9–14 July 2017, Vancouver, Canada. Lecture Notes in Computer Science. vol.10282*. Cham: Springer; 2017. p.379–395. DOI:10.1007/978-3-319-58559-8_30

8. Kolomeets M., Chechulin A. Analysis of the Malicious Bots Market. *Proceedings of the 29th Conference of Open Innovations Association, FRUCT, 12–14 May 2021, Tampere, Finland*. IEEE; 2021. p.199–205. DOI:10.23919/FRUCT52173.2021.9435421

9. Perdana R.S., Muliawati T.H., Alexandro R. Bot spammer detection in twitter using tweet similarity and time interval entropy. *Jurnal Ilmu Komputer dan Informasi.* 2015;8(1):19–25. DOI:10.21609/jiki.v8i1.280

10. *The Black Market for Social Media Manipulation*. Riga: NATO StratCom COE; 2018.

11. Chavoshi N., Hamooni H., Mueen A. DeBot: Twitter Bot Detection via Warped Correlation. *Proceedings of the 16th International Conference on Data Mining, ICDM, 12–15 December 2016, Barcelona, Spain*. IEEE; 2016. p.817–822. DOI:10.1109/ICDM.2016.0096

12. Dorri A., Abadi M., Dadfarnia M. SocialBotHunter: Botnet Detection in Twitter-Like Social Networking Services Using Semi-Supervised Collective Classification. *Proceedings of the 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech, 12–15 August 2018, Athens, Greece*. IEEE; 2018. p.496–503. DOI:10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00097

13. Vitkova L., Kotenko I., Kolomeets M., Tushkanova O., Chechulin A. Hybrid Approach for Bots Detection in Social Networks Based on Topological, Textual and Statistical Features. *Proceedings of the Fourth International Scientific Conference Intelligent Information Technologies for Industry, IITI'19, 2–7 December 2019, Ostrava – Prague, Czech Republic. Advances in Intelligent Systems and Computing. vol.1156*. Cham: Springer; 2020. p.412–421. DOI:10.1007/978-3-030-50097-9_42

14. García-Orosa B., Gamallo P., Martín-Rodilla P., Martínez-Castaño R. Hybrid Intelligence Strategies for Identifying, Classifying and Analyzing Political Bots. *Social Sciences*. 2021;10(10):357. DOI:10.3390/socsci10100357

15. Yang K.C., Hui P.M., Menczer F. Bot Electioneering Volume: Visualizing Social Bot Activity During Elections. *Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, 13–17 May 2019, San Francisco, USA*. New York: Association for Computing Machinery; 2019. p.214–217. DOI:10.1145/3308560.3316499

16. Adrian R., Kaiser J. The False positive problem of automatic bot detection in social science research. *PLoS ONE*. 2020;15(10):e0241045. DOI:10.1371/journal.pone.0241045

17. Boneh D., Grotto A.J., McDaniel P., Papernot N. How Relevant is the Turing Test in the Age of Sophisbots? *IEEE Security & Privacy*. 2019;17(6):64–71. DOI:10.1109/MSEC.2019.2934193

18. Aiello L.M., Barrat A., Schifanella R., Cattuto C., Markines B., Menczer F. Friendship prediction and homophily in social media. *ACM Transactions on the Web*. 2012;6(2):1–33. DOI:10.1145/2180861.2180866

19. Kolomeets M. *Security Datasets – MKVK2021*. URL: https://github.com/guardeec/datasets#mkvk2021 [Accessed 28th February 2023]

20. Branitskiy A., Levshun D., Krasilnikova N., Doynikova E., Kotenko I., Tishkov A., Vanchakova N., Chechulin A. Determination of Young Generation's Sensitivity to the Destructive Stimuli based on the Information in Social Networks. *Journal of Internet Services and Information Security*. 2019;9(3):1–20.

21. Pronoza A.A., Vitkova L.A., Chechulin A.A., Kotenko I.V., Saharov D.V. Methodology for disseminating information channels analysis in social networks. *Vestnik of Saint-Petersburg University applied mathematics. Computer science. Control processes.* 2018;14(4):362–377. (In Russ). DOI:10.21638/11702/spbu10.2018.409

Статья поступила в редакцию 06.02.2023; одобрена после рецензирования 14.02.2023; принята к публикации 15.02.2023.
The article was submitted 06.02.2023; approved after reviewing 14.02.2023; accepted for publication 15.02.2023.

Информация об авторах:

КОЛОМЕЕЦ Максим Вадимович — старший научный сотрудник Санкт-Петербургского Федерального исследовательского центра Российской академии наук,
https://orcid.org/0000-0002-7873-2733

ЧЕЧУЛИН Андрей Алексеевич — кандидат технических наук, ведущий научный сотрудник Санкт-Петербургского Федерального исследовательского центра Российской академии наук, доцент кафедры защищенных систем связи Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича,
https://orcid.org/0000-0001-7056-6972