К. Е. ИЗРАИЛОВ
Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН),
Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А.Бонч-Бруевича (СПбГУТ), Санкт-Петербург

П. Е. ЖУКОВСКАЯ, П. А. КУРТА
Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А.Бонч-Бруевича (СПбГУТ), Санкт-Петербург

А. А. ЧЕЧУЛИН
Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН),
Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А.Бонч-Бруевича (СПбГУТ), Санкт-Петербург

ИССЛЕДОВАНИЕ СПОСОБА ОПРЕДЕЛЕНИЯ СТОЙКОСТИ ПАРОЛЯ К ПЕРЕБОРУ НА БАЗЕ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ

Поставлена задача на разработку опросной системы, позволяющей делать предсказания относительно стойкости паролей пользователя к перебору. Предложен способ создания такой системы, основанный на методах машинного обучения, а также описана гипотеза, лежащая в его основе. Предложены шаги способа, состоящие из разработки набора вопросов, реализации самой системы, обучения и тестирования искусственной нейронной сети. Приведены результаты проведенного эксперимента, произведено их обсуждение и сделаны выводы касательно развития способа.

Введение. Задача обеспечения безопасности доступа в информационную систему является одной из основополагающих в области информационной безопасности [1]. И если программно-аппаратные средства и способны обеспечивать высокий уровень защиты, то человеческий фактор все также остается основной причиной угроз. Так, безответственное отношение пользователей к сложности своих паролей позволяет произвести их перебор злоумышленником, что приводит к несанкционированному доступу к информации. Зная о таких пользователях, можно как не допускать их к критическим информационным ресурсам, так и проводить дополнительные мероприятия по повышению «культуры информационной безопасности». Таким образом, выявление «слабых» пользователей с позиции парольных систем является актуальной задачей области. Для ее решения одним из способов может быть предлагаемый в статье, а именно – применение опросной системы, позволяющей предсказывать сложности паролей, создаваемых пользователем – их стойкости к перебору. В отличие от существующих способов, направленных больше на выявление инсайдеров [2], регулирования политик паролей, данный направлен на решение качественной задачи более высокого уровня – проверки и повышения благонадежности сотрудников, понижая тем самым влияние человеческого фактора на информационную безопасность. Основной трудностью решения является то, что требуется совместить качественно разные аспекты – социально-психологические, влияющие на задаваемые пользователем пароли [3] и вероятностно-статистические, определяющие сложность подбора пароля злоумышленником [4]. При этом, поскольку нахождение соответствия аспектов является отдельной сложной задачей, предлагается применение для этого машинного обучения в части искусственной нейронной сети (часто применяемой в информационной безопасности [5]). Последняя позволит произвести классификацию или регрессию ответов на вопросы пользователя по сложности его паролей.

Схема способа. Опишем гипотезу, лежащую в основе идеи предлагаемого способа. Предположим, имеются два пользователя парольной системы. При этом первый (например, Алиса) увлекается художественной литературой, имеет домашнее животное и романтический взгляд на жизнь. Второй же (например, Боб) полностью поглощен математическими науками, увлечен программированием и компьютерными играми, а также абсолютный прагматик. Не вдаваясь в подробности различных рассуждений, а также исходя из предварительно проведенного исследования пользователей подобного типа, можно предположить следующее. С некоторой

вероятностью пароль Алисы будет иметь отношение к области гуманитарных знаний – можно предположить, что пароль будет состоять из имен литературных героев, названий произведений и т. п.; отметим, что авторские исследования частично подтвердили подобную догадку. С другой стороны, Боб, изначально являясь более осведомленным в IT-сфере, а также понимающий алгоритмические принципы переборов (по алфавиту или словарю), предложит более сложный пароль, использующий возможно жаргонные термины или фразы, общеупотребимые лишь в его «цифровом комьюнити». Как следствие, можно предположить большую сложность подбора его пароля. Естественно, данные рассуждения являются тривиальными, частными и не могут лежать в основе построения общей логики определения сложности пароля; так, например, бесконечное усложнение пароля приведет к тому, что он не будет запомнен пользователем, а значит, окажется бесполезным. Тем не менее, суть гипотезы, вытекающей из рассуждений выше, заключается в следующем: «Социально-психологические особенности пользователя парольной системы оказывают влияние на сложность создаваемых им паролей».

Исходя из решения задачи выявления слабых пользователей, опишем схему предлагаемого способа.

Шаг 1. Необходимо составить список вопросов, который бы разносторонне отражал пользователя парольной системы. При этом делать какие-либо предположения касательно влияния ответов на свойства итогового пароля будет категорически ошибочным, поскольку как уже указывалось, такие взаимосвязи являются отдельно стоящей трудно-решаемой задачей, требующей проведения не одного научного исследования.

Шаг 2. Необходимо разработать опросную систему, которая помимо принятия ответов на вопросы запрашивала бы у пользователя типовой для него пароль. Также целесообразно запросить пароль в начале и конце опроса для проверки того, что он действительно хорошо запоминаем пользователем, а значит отражает особенности последнего.

Шаг 3. Необходимо выбрать механизм определения сложности пароля. Для этого возможно применение различных онлайн-ресурсов. Так, например, сайт Kaspersky Password Check (https://password.kaspersky.com/) выдает время подбора вводимого пароля, что может считаться его сложностью.

Шаг 4. Необходимо выбрать группу несвязанных друг с другом пользователей и произвести их тестирование с помощью разработанной системы. В результате будет получена таблица ответов на вопросы и сложности их типовых паролей.

Шаг 5. Необходимо построить и обучить искусственную нейронную сеть для решения задачи регрессии, где признаками являлись бы вопросы, входными данными – ответы на вопросы, выходными данными – сложность пароля.

Шаг 6. Необходимо протестировать обученную искусственную нейронную сеть на новых входных данных и сравнить результаты, получаемые на онлайн-ресурсе, с результатами, получаемыми сетью.

Необходимо уточнить смысл термина «типового пароля». Под последним понимается тот, который является обычным, легко запоминаемым для пользователя, отражает его собственные социально-психологические особенности, представляет собой некую базу для реально используемых. Данная гипотеза о наличии у пользователей некоторых закономерностей среди всех создаваемых паролей была обоснована проведенным авторским исследованием, предваряющим данное. Например, замечено, что люди, обладающие гуманитарным складом ума и хорошо знакомые с художественной литературой, зачастую используют в частях паролей имена известных персонажей, фразы из классических произведений, исторические события. Сотрудники научной сферы могут применять названия физических установок, значения констант, даты великих открытий. Представители же неформальной молодежи часто оперируют жаргонными выражениями, повторениями символов, заменой букв на визуально подобные знаки. Естественно, такие закономерности не являются доказанным правило; тем не менее, они позволяют утверждать, что понятие «типовой пароль» имеет право на существование. В интересах вышесказанного и предназначены проверки на Шаге 2. Так, успешное их прохождение подтвердит, что пароль

действительно является типовым для пользователя, поскольку был создан в начале опроса и безошибочно воспроизведен в конце. Естественно, система полагается на то, что пользователь запомнил придуманный пароль, а, например, не записал его на носителе информации.

А поскольку система предлагает именно придумать новый, легко запоминающийся пароль, а не ввести реально используемый в жизни, то ни при обработке его системой, ни при отправке на онлайн-ресурс (на Шаге 3) конфиденциальность пользовательской информации не будет нарушена.

В результате выполнения всех шагов будет получена оценка того, насколько успешно может применяться искусственная нейронная сеть для предсказания сложности паролей пользователей на основании их ответов на вопросы.

Результаты эксперимента. В интересах подтверждения гипотезы и проверки работоспособности предложенного способа был разработан прототип опросной системы – с использованием Google Form для проведения опроса, Kaspersky Password Check для оценки сложности пароля (отнормированной от 0 до 1) и программы на языке Python для реализации искусственной нейронной сети. Опрос состоял из 37 вопросов вида «Ваш пол?», «Есть ли домашние животные?», «Часто ли Вы фотографируете?» и др. Для упрощения реализации система запрашивала пароль только в конце опроса (т.е. один раз, а не два), не проверяя тем самым, насколько он действительно близок пользователю. Это конечно вносит некоторый шум в корректность ответов, но не является критичным.

Всего было опрошено 83 пользователя, 67 из которых использовались для обучения искусственной нейронной сети, а 16 – для тестирования. Затем, для тестируемых пользователей было произведено сравнение сложности их пароля – реальной с помощью сайта и прогнозируемой с помощью искусственной нейронной сети. Результаты сравнения показаны на рисунке.

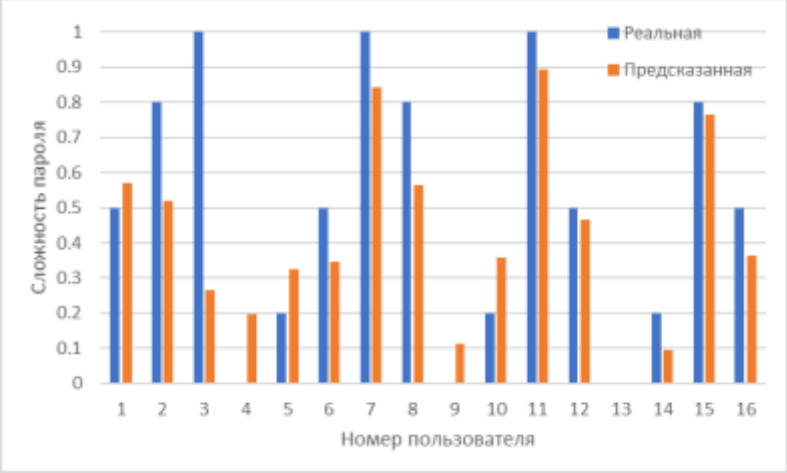

Рисунок. Результаты сравнения реальной и предсказанной сложности пароля пользователя

Согласно результатам тестирования прототипа опросной системы (рисунок), она показала хорошие результаты – средний разброс реальных и предсказываемых сложностей паролей составил ~17%.

Обсуждение. Естественно, полученные результаты не могут служить гарантированной достоверностью работоспособности способа и прототипа системы, поскольку было как допущено упрощение в последней, так и общая выборка пользователей является достаточно малой и скорее всего не имеющей требуемого качества. Тем не менее, уже сейчас можно утверждать, что определенные закономерности между социально-психологическими особенностями человека и степенью сложность создаваемых им паролей существует.

Выводы. В интересах предварительной проверки гипотезы была разработана и исследована опросная система, направленная на обнаружение слабых пользователей (с позиции парольных систем) превентивным способом – т. е. еще до их непосредственной работы. Первые результаты

показали гипотетическую работоспособность способа. Тем не менее, для оценки эффективности предложенного способа необходимо расширение проверяемых вопросов (включая касающиеся интерфейсных элементов систем [6]), получение более независимой выборки пользователей, проведение полномасштабного тестирования на их большем количестве, оценка различных метрик качества результатов. Все это планируется осуществить авторами в дальнейшей научной работе.

Исследование проводится при поддержке Минобрнауки России в рамках Соглашения № 05.607.21.0322 (идентификатор RFMEFI60719X0322).

ЛИТЕРАТУРА

1. **Буйневич М.В., Васильева И.Н., Воробьев Т.М., Гниденко И.Г., Егорова И.В. и др.** Защита информации в компьютерных системах: монография. СПб.: Санкт-Петербургский государственный экономический университет, 2017. 163 с.
2. **Буйневич М.В., Власов Д.С.** Сравнительный обзор способов выявления инсайдеров в информационных системах // Информатизация и связь. 2019. № 2. С. 83-91.
3. **Войскунский А.Е., Нафтульев А.И.** Актуальные психологические проблемы кибер-этики // Гуманитарная информатика. 2007. № 3. С. 31-39.
4. **Блинов А.С., Степаненко М.А.** Оценка достаточной сложности пароля для безопасного использования на веб-ресурсах // Исследования в области естественных наук. 2014. № 8 (32). С. 24-27.
5. **Буйневич М.В., Израилов К.Е.** Обобщенная модель статического анализа программного кода на базе машинного обучения применительно к задаче поиска уязвимостей // Информатизация и связь. 2020. № 2. С. 143-152.
6. **Ахунова Д.Г., Вострых А.В., Курта П.А.** Оценка пользовательского интерфейса информационных систем посредствам моделей качества программного обеспечения // Информатизация и связь. 2020. № 2. С. 127-135.

K.E. Izrailov (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIARAS), Saint-Petersburg; The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, (SPbSUT), Saint-Petersburg)

P.E. Zhukovskaya, P.A. Kurta (The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, (SPbSUT), Saint-Petersburg)

A.A. Chechulin (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIARAS), Saint-Petersburg; The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, (SPbSUT), Saint-Petersburg)

Research of the Method for Determining the Password Resistance to Brute Force on the Basis of an Artificial Neural Network

The task is to develop a polling system that makes it possible to make predictions about the strength of user passwords to brute force. A method for creating such a system based on machine learning methods is proposed, and the hypothesis underlying are described. There are descriptions of the method, which include the development of a set of questions, the implementation of the system itself, training and testing of an artificial neural network. The results of the experiment are given, they are discussed and conclusions are drawn regarding the development of the method.