К ВОПРОСУ О ПАРОЛЬНОЙ ЗАЩИТЕ ПОЧТОВЫХ СЕРВИСОВ

Марков Г.А.[1],Шарунов В.А.[2]

Подсистема аутентификации крайне важна в области информационной безопасности, так как является первым защитным рубежом компьютерной системы. Несмотря на неуклонное развитие механизмов информационной безопасности, наиболее используемым средством аутентификации является пароль. Основной уязвимостью такого механизма защиты считается выбор не стойкого пароля. В 2014–2015 годах произошел ряд утечек парольных баз крупных интернет–компаний, что позволило провести исследование стойкости реальных паролей. Следует констатировать, что за прошедшее время защита парольных систем не сильно продвинулась вперед, в основном видна тенденция роста требований к интерфейсу ввода пароля. При этом, до сих пор стоит вопрос, какие пароли можно считать стойкими, а какие нет. В работе приводятся примеры оценки парольных систем, а также проведен анализ утекших паролей на предмет их стойкости по разработанным требованиям. Проверка стойкости проводилась при помощи использования метрик (показателей стойкости паролей). Данные метрики являлись основой для формулирования объективных требований к стойкости парольной системы.

***Ключевые слова**: аутентификация, пароль, метрика, информационная безопасность, защита информации, парольная система*

Введение

Несмотря на то, что вопросы стойкости парольных систем подлежат перманентному исследованию, в практическом плане этот вопрос не получил завершения по субъективным причинам. Например, ряд разработчиков программных систем по-разному трактуют вопрос стойкости парольной системы, а пользователи, зачастую, не полностью соблюдают политику безопасности системы аутентификации.

За последние два года произошел ряд крупных утечек парольных баз почтовых интернет-сервисов (Яндекс, Google, Mail, Dropbox и др.). Это позволило провести исследование стойкости паролей, опираясь на неформальные и формальные показатели.

Понятие стойкости парольной защиты

Рассмотрим формулу вероятности подбора пароля [7]:

$$P = \frac{V \cdot T}{|A|^n},$$

где V – скорость подбора пароля злоумышленником, T – срок действия пароля, $|A|^n$ – мощность пространства паролей, n – длина пароля.

В соответствии с приведенной формулой можно сделать вывод, что на стойкость пароля в основном влияют частота смены пароля и мощность пространства паролей, которая характеризуется длиной и используемым алфавитом при составлении пароля.

В связи со сказанным можно сформулировать простые содержательные критерии, по которым пароль считается стойким:

- длина пароля должна быть не менее 8 символов;
- должны учитываться символы разных регистров;
- должны использоваться цифры;
- должны использоваться спецсимволы;
- основой пароля не должно быть какое-либо слово;
- пароль не должен состоять из данных, связанных с владельцем пароля.

В то же время в литературе ведется дискуссия насчет формальных требований к парольным системам [1-10].

Метрики стойкости паролей

Приведем несколько наиболее известных классов показателей стойкости парольных систем:

- численные метрики [3];
- вероятностные метрики [2, 10];
- информационная энтропия по Шеннону;
- эвристические модификации энтропии;
- вероятностные модификации энтропии [5, 6].

К численным метрикам относятся значения времени полного перебора пароля. К сожалению, такой метод не учитывает целенаправленного перебора и угадывания.

Вероятностные метрики получаются исходя из имеющийся парольной статистики для конкретных систем, что не всегда можно сделать на практике.

1 Марков Георгий Алексеевич, МГТУ им.Н.Э.Баумана, Москва, gm@cnpo.ru,
2 Шарунов Владислав Александрович, University of Greenwich, London (UK), mrvo788@gmail.com

В данной работе будут рассмотрены энтропия по Шеннону и эвристическая энтропия (рекомендованная стандартом NIST SP 800-22). Отличие методов в том, что в энтропии по Шеннону предполагается, что пароли генерируются случайным датчиком, а в случае эвристической энтропии пароль составляется человеком.

Энтропия по Шенону вычисляется следующим образом:

$$H = \log_2|A|^n = n \cdot \log_2|A| = n\frac{\ln|A|}{\ln 2},$$

где $|A|$ - мощность алфавита, n – длина пароля.

Метрика указывает на то, что чем сложнее алфавит и чем длиннее пароль, тем он более стойкий.

Приведем пример вычисления энтропии по Шеннону (табл.1).

Таблица 1.

Пример вычисления энтропии

Алфавит/длина	5	6	7	8
Латиница	23.5	28.2	32.9	37.6
Цифры	16.6	19.9	23.2	26.5
Латиница+верхний регистр+цифры	29.7	35.7	41.6	47.6
Латиница+кирилица +верхний регистр+цифры	35	41.9	48.9	55.9

Энтропию пароля по рекомендациям NIST можно вычислить по следующие [6]:

$$S = 4 + \sum_{i=2}^{8} 2 + \sum_{i=9}^{20} 1.5 + \sum_{i=21}^{n} 1 + 6\chi_A$$

где $i \le n$, n – длина пароля, χ_A – характеристическая функция наличия в пароле неалфавитных символов или символов верхнего регистра.

Данную формулу можно описать следующим образом: первый символ пароля получает значение 4 бит, каждый имеющийся далее символ со второго по восьмой получают по 2 бита, с 9-го по 20-ый по 1.5 бита и каждый последующий по одному биту. При наличии неалфавитных символов или символов верхнего регистра к полученному результату прибавляется 6 бит.

По данным метрикам будем считать, что пароль стойкий, если он соответствует энтропии [6]:

- по Шеннону - 56 бит и более,
- по рекомендациям NIST - 24 и более бит.

Следует наложить ограничение на указанные критерии: если пароль зафиксирован в базах для подбора паролей (словарях), то энтропия сводится к нулю.

Результаты исследования

При помощи исследовательской программы были обработаны несколько парольных баз, выложенных хакерами в открытый доступ в глобальной сети интернет в прошлом году. По каждой парольной базе была получена определенная статистика, представленная ниже. Результаты исследования скомпрометированной базы паролей Яндекс (1 261 809 паролей) представлены в табл. 2-4, Mail.ru (45 000) – табл. 5-7, Google (4 926 673) – табл. 8-10.

Таблица 2.

Длина паролей (Яндекс)

Длина пароля	Количество паролей
6	380732
7	174782
8	282641
9	130676
10	103926
11	71948
12	45387
13	20127
14	14950
15	9895
16	7646
17	3487
18	3104
19	1747
20	2660

Таблица 3.

Топ-10 повторяющихся паролей (Яндекс)

Пароль	Количество повторений
123456	39177
123456789	13892
111111	9826
qwerty	7926
1234567890	5853
1234567	4668
7777777	4606
123321	4324
000000	3304
123123	3031

Таблица 4.
Алфавит паролей (Яндекс)

Используемый алфавит	Число паролей
Пароли, состоящие только из цифр	608125
Пароли, состоящие из символов	233561
Пароли, состоящие только из нижнего регистра	218319
Только верхний регистр	3136
Похожие на номер мобильного телефона	40980
Совпадение с логином	1489
Похожие на даты	171906
Подходящие под содержательное описание стойкого пароля	345
Подходящие по стойкости по Шенону	143802
Подходящие по стойкости по NIST	108951

Таблица 5.
Длина паролей (Mail.ru)

Длина пароля	Количество паролей
6	17484
7	4155
8	12562
9	3212
10	2421
11	1399
12	1106
13	627
14	438
15	293
16	205
17	12
18	20
19	2
20	15

Таблица 6.
Топ-10 повторяющихся паролей (Mail.ru)

Пароль	Количество повторений
qwerty	4291
987654321	1385
4815162342	661
11111111	615
123123123	578
789456123	448
12341234	408
147852369	380
444444	353
q1w2e3	331

Таблица 7.
Алфавит паролей (Mail.ru)

Используемый алфавит	Число паролей
Пароли, состоящие только из цифр	18806
Пароли, состоящие из символов	14650
Пароли, состоящие только из нижнего регистра	13835
Только верхний регистр	53
Похожие на номер сотового телефона	138
Совпадение с логином	3619
Похожие на даты	9287
Подходящие под содержательное описание стойкого пароля	5
Подходящие по стойкости по Шеннону	3916
Подходящие по стойкости по NIST	3274

Таблица 8.
Длина паролей (Google)

Длина пароля	Количество паролей
6	924154
7	663510
8	1422999
9	683315
10	682811
11	152256
12	93202
13	42387
14	24853
15	14851
16	7291
17	2549
18	1781
19	1082
20	1166

Таблица 9.
Топ-10 повторений паролей (Google)

Пароль	Количество повторений
123456	47918
password	11554
123456789	11160
12345	8096
querty	5918
12345678	5250
111111	3521
abc123	3011
123123	2972
1234567	2911

Таблица 10.

Алфавит паролей (Google)

Используемый алфавит	Число паролей
Пароли, состоящие только из цифр	774669
Пароли, состоящие из символов	1968873
Пароли, состоящие только из нижнего регистра	1968873
Только верхний регистр	0
Похожие на номер сотового телефона	22751
Совпадение с логином	45010
Похожие на даты	156142
Подходящие под содержательное описание стойкого пароля	0
Подходящие по стойкости по Шеннону	290530
Подходящие по стойкости по NIST	157475

Выводы

Сравнительный анализ полученной статистики с известной ранее [11] показал тенденцию незначительного усиления парольной защиты. Это связано с тем, что, ряд интернет-сервисов определил более строгие правила к соответствующим интерфейсам, например, усилил требование к длине паролей (не менее 6 символов) и использованию относительно сложного алфавита. Однако - о чем свидетельствует статистика - указанное не останавливает неорганизованных и беспечных пользователей в выборе легко подбираемых паролей, и число топ-500 паролей практически не меняется из года в год [11].

В целом проведенное исследование подтвердило, что система аутентификации остается весьма уязвимой (только 10 % паролей модно считать надежными), что обуславливает создание интегрированных систем защиты информации и развитие систем менеджмента информационной безопасности.

В заключение следует отметить, что использование энтропийных метрик вместо вербальных описаний более практично при определении технических требований к системам обеспечения безопасности информации, т.к. они легче поддаются автоматизации и контролю. Кроме того, использование формальных показателей позволяет снизить степень субъективизма, присутствующую при анализе безопасности систем.

Рецензент: кандидат технических наук Цирлов Валентин Леонидович, v.tsirlov@bmstu.ru

Литература:

1. Беленко А. Пароли: стойкость, политика назначения и аудит // Защита информации. Инсайд. 2009. № 1 (25). С. 61-64.
2. Гуфан К.Ю., Новосядлый В.А., Эдель Д.А. Оценка стойкости парольных фраз к методам подбора // Открытое образование. 2011. №2. 127-130 с.
3. Евтеев Д. Анализ проблем парольной защиты в российских компаниях. 2009. 33 с. URL: http://www.ptsecurity.ru/download/PT-Metrics-Passwords-2009.pdf (дата обращения: 08.12.2015).
4. Заркумова Р.Н. Исследование количественных характеристик системы парольной защиты информации // Сборник научных трудов НГТУ. 2010. № 2(60). С.83-88.
5. Марков Г.А. К вопросу об определении стойкости парольных систем // Сборник трудов Третьей всероссийской НТК «Безопасные информационные технологии» / под. Ред. В.А.Матвеева. М: НИИ РЛ МГТУ им.Н.Э.Баумана. 2012. С.21-23.
6. Марков Г.А. Метрики стойкости парольной защиты // Молодежный научно-технический вестник. 2013. № 2. С. 28.
7. Методы оценки несоответствия средств защиты информации/А.С.Марков, В.Л.Цирлов, А.В.Барабанов. М.: Радио и связь, 2012. 192 с.
8. Тюрин К.А., Сёмин Р.В. Анализ стойкости парольных фраз на основе информационной энтропии // Известия ЮФУ. Технические науки. 2015. № 5 (166). С. 18-27.
9. Шибанов С.В., Карпушин Д.А. Сравнительный анализ современных методов аутентификации пользователя // Математическое и программное обеспечение систем в промышленной и социальной сферах. 2015. № 1 (6). С. 33-37.
10. Bonneau J. Guessing human-chosen secrets // Technical Report UCAM-CL-TR-819. 2012. 161 p.
11. The Top 500 Worst Passwords of All Time, 2008. URL: http://www.whatsmypass.com/the-top-500-worst-passwords-of-all-time

ABOUT INFORMATION SECURITY OF EMAIL SERVICES

Markov G.A.[3], Sharunov V.A.[4]

In 2014–2015 there was a series of leaks of password databases of large companies such as Yandex, Google, Mail, Dropbox. Since then, the protection of password systems are not much has moved forward, mostly visible upward trend password requirements. For example, most sites are introduced requirements for the minimum length and contents of the alphabet in passwords. Therefore, until now, the question is which passwords can be considered persistent, and which are not. The paper provides examples of assessment of password systems, as well as an analysis of leaked passwords for their resistance by the above requirements.

***Keywords**: password, metric, information security, data protection, password system*

Reference:

1. Belenko A. Paroli: stoykost', politika naznacheniya i audit, Zashchita informatsii. Insayd. 2009. No 1 (25), pp. 61-64.
2. Gufan K.Yu., Novosyadlyy V.A., Edel' D.A. Otsenka stoykosti parol'nykh fraz k metodam podbora, Otkrytoe obrazovanie. 2011. No 2, pp. 127-130.
3. Evteev D. Analiz problem parol'noy zashchity v rossiyskikh kompaniyakh. 2009. 33 P. URL: http://www.ptsecurity.ru/download/PT-Metrics-Passwords-2009.pdf.
4. Zarkumova R.N. Issledovanie kolichestvennykh kharakteristik sistemy parol'noy zashchity informatsii, Sbornik nauchnykh trudov NGTU. 2010. No 2(60), pp.83-88.
5. Markov G.A. K voprosu ob opredelenii stoykosti parol'nykh system, Sbornik trudov Tret'yey vserossiyskoy NTK «Bezopasnye informatsionnye tekhnologii», pod. Red. V.A.Matveeva. M: NII RL MGTU im.N.E.Baumana. 2012, pp.21-23.
6. Markov G.A. Metriki stoykosti parol'noy zashchity, Molodezhnyy nauchno-tekhnicheskiy vestnik. 2013. No 2, pp. 28.
7. Metody otsenki nesootvetstviya sredstv zashchity informatsii, A.S.Markov, V.L.Tsirlov, A.V.Barabanov. M.: Radio i svyaz', 2012. 192 P.
8. Tyurin K.A., Semin R.V. Analiz stoykosti parol'nykh fraz na osnove informatsionnoy entropii // Izvestiya YuFU. Tekhnicheskie nauki. 2015. No 5 (166), pp. 18-27.
9. Shibanov S.V., Karpushin D.A. Sravnitel'nyy analiz sovremennykh metodov autentifikatsii pol'zovatelya // Matematicheskoe i programmnoe obespechenie sistem v promyshlennoy i sotsial'noy sferakh. 2015. No 1 (6), pp. 33-37.
10. Bonneau J. Guessing human-chosen secrets // Technical Report UCAM-CL-TR-819. 2012. 161 p.
11. The Top 500 Worst Passwords of All Time, 2008. URL: http://www.whatsmypass.com/the-top-500-worst-passwords-of-all-time

3 Georii Markov, Bauman MSTU, Moscow, gm@cnpo.ru,
4 Vladislav Sharunov, University of Greenwich, London (UK), mrvo788@gmail.com