На рисунке 2 показан процесс отправки данных, а на рисунке 3 – процесс получения данных, обеспечивающие доставку и целостность данных.

Авторы считают, что в данной работе новыми являются следующие положения и результаты: предложен алгоритм обеспечения доставки и целостности данных между агентами в СЭД, выявлены требования к системе СЭД, разработана архитектура СЭД на основе технологии удостоверяющих центров и интеллектуальных агентов.

Применение полученных результатов возможно в различных отраслях. В данный момент происходит разработка версии системы для факультета университета и процесса управления ремонтными работами.

Литература

1. *Нгуен Д. Х., Кизим А. В., Камаев В. А.* Применение многоагентной системы для обеспечения безопасности инфраструктуры открытых ключей // Нечёткие системы и мягкие вычисления (НСМВ-2009): сб. ст. 3-й всерос. науч. конф., 21-24 сент. 2009 г. / ВолгГТУ [и др.]. - Волгоград, 2009. - Т. 1. - С. 172-179.

2. Переход от бумажного к смешанному документообороту [Электронный ресурс]. – [2009] . – Режим доступа: http://www.hr-portal.ru/article/perekhod-ot-bumazhnogo-k-smeshannomu-dokumentooborotu

3. *Rao M. P.* Georgeff. "BDI-agents: From Theory to Practice". [Электронный ресурс]. – [1995] . – Режим доступа: https://www.aaai.org/Papers/ICMAS/1995/ICMAS95-042.pdf

4. Протокол управления передачей. Программная спецификация протокола DARPA INTERNET. [Электронный ресурс]. – [2006] . – Режим доступа: http://www.protocols.ru/files/RFC/rfc793.pdf

5. *Нгуен Д. Х., Кизим А. В.* Вопросы защиты информации в сети // Прикаспийский журнал: управление и высокие технологии. - 2008. - № 2. - С. 13-16.

УДК 004.056.53

ОЦЕНКА СТОЙКОСТИ ПАРОЛЬНЫХ ФРАЗ К МЕТОДАМ ПОДБОРА

***Ю. К. Гуфан**, зам. директора*
Тел.:(863) 201-28-24, (863) 201-28-17, e-mail: sva@sfedu.ru
***В. А. Новосядлый**, к.. ф.-мат. н, зав. лабораторией*
Тел.:(863) 201-28-24, (863) 201-28-17, e-mail: sva@sfedu.ru
***Д. А. Эдель**, научный сотрудник*
Тел.:(863) 201-28-24, (863) 201-28-17, e-mail: sva@sfedu.ru
ФГНУ НИИ «Спецвузавтоматика»

Various methods of password strength estimation based on information entropy are described.The conclusions about popular users passwords protection are made. The example of statistical attack is given.

В работе рассматриваются различные методы оценки стойкости парольных фраз на основе подсчета информационной энтропии. Делаются выводы о защищенности популярных паролей пользователей. Приводится пример атаки, учитывающей статистики символов в паролях.

Ключевые слова: информационная энтропия, пароль, атака грубой-силы.

Keywords: information entropy, password, brute-force attack.

Введение

В основе данной работы лежат исследования паролей пользователей сети Интернет. Одним из отличительных свойств таких паролей является их не случайность, связанная с необходимостью их легкого запоминания. Не случайность ключей не учитывается создателями криптографических алгоритмов и тем самым такие пароли подчас являются слабым звеном в системе безопасности. При этом пользователи, желая увеличить уровень своей защищенности, могут использовать различные рекомендации по составлению своих паролей.

Общепринятые рекомендации [1, 2] по составлению безопасных паролей можно сформулировать следующим образом: пароль является безопасным, если он достаточно длинный и в нем

используется символы из различных групп (маленькие и большие буквы, цифры, специальные символы).

Таким образом, согласно общим рекомендациям:

1)«*aaaaaa*» (шесть символов) оценивается как более безопасный чем «*rxgnp*» (пять символов);

2)«*aaa111*» (буквы и цифры) более безопасен чем «*LKXRWB*» (буквы);

3)«*aaa123*» (буквы и цифры) более безопасен чем "*#*&!$%*" (только специальные символы).

Во всех приведенных примерах, множество возможных паролей необходимое для подбора первого из пары больше множества которое необходимо перебрать, чтобы найти второй пароль. На основании этого свойства и делается вывод о том, что первые пароли более устойчивы к атакам нежели вторые.

Как будет показано далее, такие оценки не в полной мере отражают стойкость ключевых фраз к различным вариантам подбора, в том числе, часто более "безопасные" пароли, составленные по таким рекомендациям являются менее случайными и тем самым, более предсказуемыми для злоумышленника.

Статистический подбор пароля

Предположим, что злоумышленник пытается подобрать пароль обладая словарем популярных ключевых фраз. Проверка пароля на совпадение с записью в таком словаре является достаточно популярным методом атаки. Такой метод не позволяет злоумышленнику вскрыть любой пароль, так как в процессе подбора проверяются не все возможные варианты. Очевидно, что приведенные выше в качестве примеров «сильные», а не «слабые» пароли уязвимы для атаки по словарю. Но знание популярных паролей дает возможность злоумышленнику не только подбирать словарные ключи, но и "угадывать" пароли похожие на популярные.

Широко известен факт сохранения статистик использования символов и их групп в естественных языках [3]. Аналогичный факт имеет место и для паролей [4], не смотря на то, что в них часто используются не только слова но и годы рождения, особенности расположения символов на клавиатуре и т.д. Предположим, что злоумышленник использует вероятности встречи символов (для символа «a» обозначим $P_{"a"}$), биграмм (для биграммы «ab» – $P_{"ab"}$), следование одного символа после другого («b» после «a» – $P_{"b"|"a"}$).

Для оценки стойкости ключевых фраз от атак подбора оценим показатели их информационной энтропии. Низкая средняя энтропия символа для паролей означает, что они являются достаточно неслучайными, следовательно злоумышленник может их подобрать как похожие друг на друга.

Экспериментальная оценка

В качестве исходных данных для экспериментальной оценки стойкости ключевых фраз от методов подбора были использованы реальные пароли пользователей социальной сети «VKonakte» опубликованные в открытом доступе в сети Интернет в августе 2010 г (в общей сложности база из 125841 паролей).

Условно базу анализируемых паролей можно разбить на группы по алфавитам использованных в них символов. Для анализа будем использовать наиболее представительные группы. В ячейках таблицы 1 приведено количество паролей относящихся к каждой такой группе. Так, в 2875 паролях длины от 5 до 7 символов использовались и английские буквы, и цифры.

Таблица 1.

Количество паролей использованных для анализа по группам

	[0-9]	[a-z]	[а-я]	[a-z0-9]	[а-я0-9]
Всего записей	21923	24472	9629	41699	10059
Уникальных	7673	7615	2906	14454	3580
Длины 2-4	98	106	44	7	0
Длины 5-7	4157	3136	1149	2875	679
Длины 8-11	3199	3336	1192	7032	1716
Длины 12-16	179	841	369	3745	894

Оценить информационную энтропию можно различным образом [5], если в общем случае энтропия:

$$H(X) = -\sum_{i=1}^{N} p(x_i)\log_2(p(x_i))$$

где N - длина алфавита, $p(x_i)$ - вероятность появления i-го символа в пароле X длины L, то предположение о равновероятности появления символов дает выражение для энтропии паролей:

$$H(X) = L\frac{\log(N)}{\log(2)} \quad (1)$$

Более точно оценить энтропию символов паролей является учет вероятностей $P_{"x_i"}$, $P_{"x_i x_{i+1}"}$ и $P_{"x_{i+1}"|"x_i"}$:

$$P_{"a"} = \frac{N_{"a"}}{N_{total}},\ P_{"ab"} = \frac{N_{"ab"}}{N^2_{total}},\ P_{"b"|"a"} = \frac{N_{"a"}}{N_{total}}\frac{N_{"b"|"a"}}{N_{total}} \quad (2)$$

где N - количество встреч: $N_{"a"}$ - символов "a", $N_{"ab"}$ - биграмм "ab", $N_{"a"|"b"}$ - символов "b" после "a", а N_{total} - символов в проанализированной группе паролей.

Оценка стойкости

Приведем значения энтропии (количество бит на символ) паролей для групп из таблицы 1.

Таблица 2.

Информационная энтропия паролей

	[0-9]			[a-z]			[а-я]			[a-z0-9]			[а-я0-9]		
P=1/N	3.32			4.70			5.93			5.17			6.11		
L:[2;4]	3.05	5.40	5.48	4.26	6.63	6.66	4.53	6.17	6.03	3.81	4.16	3.74			
[5;7]	3.29	6.49	6.50	4.52	8.45	8.44	4.71	8.17	8.14	4.88	9.06	9.03	4.93	8.64	8.65
[8;11]	3.24	6.34	6.35	4.50	8.49	8.48	4.61	8.14	8.13	4.87	9.03	9.01	4.95	8.82	8.82
[12;16]	3.22	6.30	6.30	4.56	8.68	8.68	4.64	8.13	8.14	4.95	9.25	9.24	4.87	8.59	8.60

В первой строке таблице указаны значения энтропии, полученные по формуле (1). Каждая ячейка из последующих строк разбита на три части, в соответствии со значениями энтропии полученными из соображений вероятностей по формулам (2). В строках, как и ранее, объединены пароли по их длине.

Очевидно, что во всех случаях учет вероятности появления символов дает значение энтропии меньшее, чем в исходном предположении равновероятности символов. При этом, для длины паролей от 5 до 7 знаков энтропия на символ цифрового пароля при подсчете статистик символов 3.29, а для длины от 8 до 11 – 3.24. Это означает, что более длинные пароли являются менее случайными нежели короткие.

Для дальнейших выводов нам необходимо произвести подсчет относительного изменения энтропии при увеличении размера алфавита. Иными словами, если по исходному предположению энтропия при увеличении алфавита пароля растет как $\log(N^1)/\log(N^2)$ (N^1 - длина первого алфавита, N^2 - длина второго алфавита), то энтропия с учетом статистик так же должна расти в аналогичной пропорции. В таблице 3 приводятся соответствующие значения относительного изменения энтропии символов паролей:

$$K = \frac{\log(N^1)\sum_{i=1}^{N^2} p_i^2\log_2(p_i^2)}{\log(N^2)\sum_{i=1}^{N^1} p_i^1\log(p_i^1)},$$

где p^1-рассчитывается из формул (2) по алфавиту длины N^1, а p^2-по алфавиту длины N^2. В этом выражении если K меньше единицы, то пароли относящиеся ко второму алфавиту "более" не случайны относительно паролей из первого алфавита.

Таблица 3.

Относительное изменение энтропии паролей

	[0-9]|[a-z]			[a-z]|[a-z0-9]			[а-я]|[а-я0-9]		
$\log(N^1)/\log(N^2)$	0.71			0.91			0.97		
L:[2;4]	0.99	0.87	0.85	0.81	0.57	0.51			
[5;7]	0.97	0.92	0.92	0.98	0.97	0.97	0.98	0.99	0.99
[8;11]	0.98	0.95	0.94	0.98	0.97	0.97	1.00	1.00	1.00
[12;16]	1.00	0.97	0.97	0.99	0.97	0.96	0.97	0.98	0.98

Как видно из таблицы, с увеличением алфавита пользователи начинают использовать более неслучайные (более предсказуемые) пароли. Для доказательства возможности использования статистик символов злоумышленниками приведем пример подобной атаки подбора.

Пример атаки подбора с учетом статистик

Итак, предполагаем, что нам известно распределение случайных величин вероятностей символов, биграмм и триграмм:X, X^2, X^3. Выберем некоторое значение P, 0<P<1, которое будем называть *пороговым значением.* Упорядочим значения случайной величины X по убыванию их вероятностей и возьмем наименьшее число элементов, начиная с первого в выбранном упорядочении и идущих подряд так, чтобы сумма их вероятностей была больше P. В качестве первого символа предполагаемого пароля будем брать только выбранные символы.

Предположим, что выбран первый символ предполагаемого пароля. Тогда в качестве распределения вероятностей вторых символов следует взять их условные вероятности. Обозначим уже выбранный первый символ через *a,* выбираемый второй символ через *b*. Мы имеем вероятностное распределение биграмм, то есть для любой биграммы "*ab*" известна ее вероятность $P_{"ab"}$. Как и ранее, через $P_{"b"|"a"}$ обозначим условную вероятность того, что второй символ будет равен *b* в предположении, что первый символ равен *a*. Тогда указанная условная вероятность может быть получена из теоремы о вероятности произведения событий (подсчитана как показано в (2)).

Упорядочим символы алфавита по убыванию условных вероятностей. Выберем наименьшее число элементов, начиная с первого в выбранном упорядочении и идущих подряд так, чтобы сумма их условных вероятностей была больше *p*. В качестве второго символа предполагаемого пароля будем брать только выбранные символы.

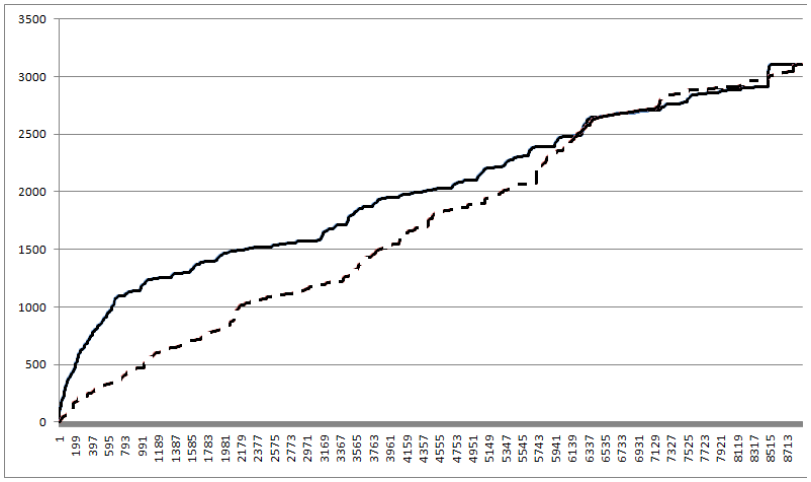

Рис. 1. Количество найденных паролей

Для третьего символа пароля *c* множество возможных вариантов для перебора с порогом *p* будет определяться из условных вероятностей $P_{"c"|"b"}$. Очевидно, что при переборе с вероятностями при значении порогов допустимых значений $[0;p] \cup (p;1]$ мы переберем все множество паролей.

На рисунке 1 приведена зависимость количества найденных паролей состоящих из английских букв длины 6 на различных шагах атаки подбора (единица шкалы – 70000 итераций) при использование методов простой грубой силы (штриховкой) и рассмотренного выше в качестве примера (сплошная линия).

Выводы

Авторы считают, что в данной работе новыми являются следующие положения и результаты:

1.Проведена оценка стойкости парольных фраз при помощи подсчета информационной энтропии различными методами, показавшая, что по статистике увеличение длины созданного пользователем пароля и алфавита использованных в нем символов приводит к уменьшению случайности пароля.

2.Приведен алгоритм атаки подбором позволяющий учитывать статистики символов, что заметно сокращает время подбора паролей большинства пользователей.

Литература

1. Автоматизированные системы. Защита от несанкционированного доступа к информации. Классификация автоматизированных систем и требования по защите информации // Руководящий документ. ФСТЭК России. 30 марта 1992.

2. Electronic Authentication Guideline. NIST. Retrieved March 27, 2008.

3. *A. Sinkov* Elementary cryptanalysis :a mathematical approach, Mathematical Association of America, Washington, D.C., 1966, 222pp.

4. *Гуфан К.Ю., Новосядлый В.А., Эдель Д.А.* О методах оценки стойкости парольных фраз // Материалы XIX научно-технической конференции «Методы и технические средства обеспечения безопасности информации» 5-10 июля 2010 г. – СПб.: Изд-во Политехн. ун-та, 2010. С. 73–74.

5. *C. E. Shannon* Prediction and entropy of printed English // Bell Systems Technical Journal. — 1951. — Т. 30. — С. 50—64.