УДК 004.056.53 Вестник УрФО № 1(27) / 2018, с. 14–18

Баранкова И. И., Михайлова У. В., Лукьянов Г. И.

РАЗРАБОТКА СИСТЕМЫ WORDSEARCH ДЛЯ ЗАЩИТЫ КОНФИДЕНЦИАЛЬНОЙ ИНФОРМАЦИИ ОТ УТЕЧКИ

Конфиденциальность информации организации или предприятия является важной составляющей, и нарушение ее может нанести значительный ущерб. На сегодняшний день существует большое количество способов и методов борьбы с утечками конфиденциальной информации. Одним из возможных и наиболее эффективных способов защиты информации это внедрение системы защиты от утечек конфиденциальных данных Data Leak Prevention (DLP). В статье приведены этапы разработки программы и результаты анализа информационного потока в корпоративной сети. Произведена оценка работы поисковых алгоритмов подстроки в строках на качественные и временные показатели. Благодаря сравнительному анализу определены слабые стороны используемых алгоритмов, а так выявлены способы обхода системы поиска. Разработаны методика использования базы конфиденциальных данных и алгоритм подстрочного поиска в электронных документах. Внедрение методики обработки базы искомых данных перед запуском поточного сканирования значительно повысило качественную характеристику системы поиска и при этом незначительно увеличило время выявления инцидентов. Произведена модернизация модуля агента с целью блокировки дальнейших манипуляций с информацией и персональной машиной при обнаружении инцидентов. Благодаря данным доработкам получилось значительно повысить эффективность разработанной системы, а именно снизить уровень утечки конфиденциальной информации.

Ключевые слова: *корпоративна сеть, конфиденциальная информация, утечка информации, DLP система, поиск подстрок в строке, WordSearch, блокировка персональной машины.*

DEVELOPMENT OF THE WORDSEARCH SYSTEM TO PROTECT FROM CONFIDENTIAL INFORMATION LEAKAGE

Confidentiality of information of the organization or the enterprise is an important component, and its violation can cause considerable damage. To date, there are a large number of ways and methods to combat the leakage of confidential information. One of the possible and most effective ways to protect information is the introduction of a system to protect against data leak Prevention (DLP). The article presents the stages of the program development and the results of the analysis of the information flow in the corporate network. The work of search algorithms of substring in lines on qualitative and temporal indexes is estimated. Due to the comparative analysis the weaknesses of the used algorithms are determined, as well as the ways to bypass the search system are revealed. The method of using the confidential data base and the algorithm of string search in electronic documents are developed. The introduction of the technique of processing the database of the required data before the launch of line scanning significantly improved the quality of the search system and thus slightly increased the time of incident detection. Upgrading module agent to block further manipulation of the information and personal machine upon detection of incidents. Thanks to these improvements it was possible to significantly improve the efficiency of the developed system, namely to reduce the level of leakage of confidential information.

Keywords: *corporate network, confidential information, information leak, DLP system, search for substrings in the string, WordSearch, lock the personal machine.*

В настоящее время один из возможных и наиболее эффективных способов мониторинга информационного потока предприятия и методов борьбы с действиями злоумышлиников являются системы защиты от утечек конфиденциальных данных Data Leak Prevention [1; 2]. Под DLP системой понимают технологии предотвращения утечек конфиденциальной информации из информационной системы, а также программные или программно-аппаратные комплексы для предотвращения различных видов утечек информации [3; 4].

Обнаружение и блокировка передачи информации из корпоративной системы в сеть осуществляется путем применения ряда стандартных функций, а именно:

– фильтрация интернет-трафика, иных информационных потоков;

– анализ контента по предварительно установленным ключевым словам, определенным выражениям, «оцифрованным» документам, учитывая совокупность всех обстоятельств [5].

Разработка системы WordSearch

Обнаружение конфиденциальной информации в DLP системах реализуется за счет применения совокупности методов поиска слов и словосочетаний по словарю и синтаксического разбора текстовых строк по формализованному шаблону (WordSeach).

Основные алгоритмы WordSeach используемые во многих DLPсистемах:

1) линейный поиск;

2) поиск Д. Кнута, Д. Мориса и В. Пратта (КМП – поиск);

3) Поиск Р. Бойера и Д. Мура (БМ-поиск);

4) Нечеткий поиск в тексте.

На основе первых трех, приведенных выше, алгоритмов поиска разработано приложение, которое производит поиск подстроки в представленном тексте (рис. 1). С использованием данного приложения про-

изведен сравнительный анализ алгоритмов поиска.

При реализации простого поиска (брут) были осуществлены доработки использования базы.

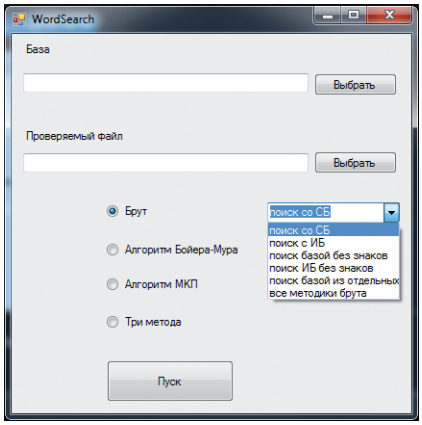

Рис. 1. Стартовое окно разработанного приложения для поиска подстроки в тексте

В результате тестирования каждого метода все они показали достаточно высокие результаты поиска. Из-за отсутствия разнородной интерпретации базы искомой информации методы КМП и БМ показали результаты ниже чем у метода «Брута». После доработки КМП и БМ методов, произведена оценка временного показателя каждого алгоритма. В качестве тестового текста использовали текст длинной 10002 символа. Характеристики стенда, на котором проводилось тестирование: CPUAthlonx4 640, ОЗУ 4Gb, Windows 10(32-bit) Pro. Результаты тестирования представлены в виде таблицы.

После доработки методов алгоритм Бойера – Мура показал наилучшее время выполнения поставленной задачи поиска.

Время работы алгоритмов поиска при различной длине искомой подстроки

Алгоритм	Время выполнения(мс)		
	Длина ≤ 10	Длина ≤ 100	Длина ≤ 250
Брут	15	93	234
КМП	5	30	50
БМ	31	31	32

Производительность всех исследуемых алгоритмов поиска представлена в графическом виде на рис. 2.

В связи с множественными поправками по базе искомых строк из-за искажения исследуемого текста, такого как замена букв,

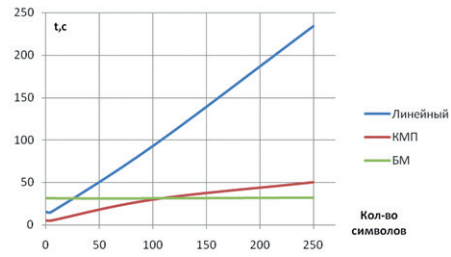

Рис. 2. Производительность алгоритмов поиска

перестановка и разрыв слов и т. п., наиболее эффективным и быстродействующий методом показал себя метод нечеткого поиска. Алгоритмы нечеткого поиска применяются в текстовых редакторах для проверки орфографии и во всем известных поисковых системах. Примером работы такой функции является вывод сообщения пользователю при запросе в поисковой системе «Возможно, вы имели ввиду…». Но как и остальные алгоритмы этот алгоритм можно обойти, поэтому необходима доработка модулей обработки сканирования перед полным внедрение алгоритма нечеткого поиска в систему.

В разработанном приложении реализована блокировка персонального компьютера сотрудника в случае обнаружения совпадений. При обнаружении искомых файлов будет произведено отключение сетевых подключений, а так же блокировка мышки и клавиатуры. При этом на экране вызовется окно с предупреждением о найденных совпадениях (рис. 3).

Рис. 3. Уведомление о найденных совпадениях

Данная блокировка снимается персоналом ответственным за работу данного ПО (сотрудники отдела безопасности). Для снятия блокировки необходимо указать в программе, какой комбинацией клавиш она будет производиться.

Заключение

С использованием разработанного нами приложения произведена оценка работы ал-

горитмов поиска подстрок в строке. При этом выявлено, что зная алгоритм каждого метода, злоумышленник может обойти DLP-систему. В результате проведенного анализа сделаны следующие выводы:

1. Необходимы разработка и создание нового алгоритма поиска с оптимальной скоростью работы, а так же учетом недостатков уже известных алгоритмов.

2. DLP система не является достаточной защитой от утечки информации, поэтому необходимо использование комплекса систем и организационных мер.

Литература

1. Баранкова И. И., Михайлова У. В., Лукьянов Г. И. DLP система: защита от утечки информации. Анализ поиска WordSearch // Актуальные проблемы современной науки, техники и образования. – 2016. – Т. 1. – № 1. С. 187–191.

2. Баранкова И. И., Михайлова У. В., Лукьянов Г. И. Прогнозирование локальных и внешних угроз на информационные серверы предприятия // Актуальные проблемы современной науки, техники и образования. – 2017. – Т. 1. – С. 217–220.

3. Баранкова И. И., Михайлова У. В., Самохвал В. Д., Огонесян Ш. У. Анализ информационных угроз ВУЗА // Актуальные проблемы современной науки, техники и образования. – 2013. – Т. 2. – № 71. – С. 157–159.

4. Коновалов М. В., Михайлова У. В., Хусаинов А. А., Санарбаев Р. Ж. Алгоритмы шифрования данных // Актуальные проблемы современной науки, техники и образования. – 2013. – Т. 2. – № 71. – С. 159–161.

5. Михайлова У. В., Коновалов М. В., Гуринец К., Кучербаева Э. Ф. Идентификация личности // Актуальные проблемы современной науки, техники и образования. – 2013. – Т. 2. – № 71. – С. 164–166.

References

1. Barankova I.I, Mikhailova U.V., Lukyanov G.I. DLP sistema: zashchita ot utechki informatsii. Analiz poiska WordSearch[DLP system: protection against information leakage. WordSearch search analysis] // Aktual'nye problemy sovremennoy nauki, tekhniki i obrazovaniya [Actual Problems of Modern Science, Technology and Education], 2016. T. 1. № 1. P. 187-191.

2. Barankova I.I, Mikhailova U.V., Lukyanov G.I. Prognozirovanie lokal'nykh i vneshnikh ugroz na informatsionnye servery predpriyatiya [Forecasting of local and external threats to enterprise information servers] // Aktual'nye problemy sovremennoy nauki, tekhniki i obrazovaniya [Actual Problems of Modern Science, Technology and Education], 2017. T. 1. P. 217-220.

3. Barankova I.I, Mikhailova U.V., Samohval V.D., Oganesyan Sh.U. Analiz informatsionnykh ugroz VUZA [Analysis of information threats of the University] // Aktual'nye problemy sovremennoy nauki, tekhniki i obrazovaniya [Actual Problems of Modern Science, Technology and Education], 2013. T. 2. № 71 P. 157-159.

4. Konovalov M.V., Mikhailova U.V., Husainov A.A., Sanarbaev R.J. Algoritmy shifrovaniya dannykh [Data Encryption Algorithms] // Aktual'nye problemy sovremennoy nauki, tekhniki i obrazovaniya [Actual Problems of Modern Science, Technology and Education], 2013. T. 2. № 71 P. 159-161.

5. Mikhailova U.V., Konovalov M.V., Gurinec K., Kucherbaeva E.F. Identifikatsiya lichnosti [Identification of a person] // Aktual'nye problemy sovremennoy nauki, tekhniki i obrazovaniya [Actual Problems of Modern Science, Technology and Education], 2013. T. 2. № 71 P. 164-166.

БАРАНКОВА Инна Ильинична, доктор технических наук, заведующий кафедрой ИиИБ, Магнитогорский государственный технический университет им. Г. И. Носова. 455000, г. Магнитогорск, пр. Ленина 38. E-mail: inna_barankova@mail.ru

МИХАЙЛОВА Ульяна Владимировна, кандидат технических наук, доцент кафедры ИиИБ, Магнитогорский государственный технический университет им. Г. И. Носова. 455000, г. Магнитогорск, пр. Ленина 38. E-mail: ylianapost@gmail.com

ЛУКЬЯНОВ Георгий Игоревич, ассистент кафедры ИиИБ Магнитогорский государственный технический университет им. Г.И. Носова 455000, г. Магнитогорск, пр. Ленина 38. E-mail: decorsi@mail.ru.

BARANKOVA Inna, Department, Nosov Magnitogorsk State Technical University (NMSTU), D.Sc.,Head of Computer Science and Information Safety Engineering (CSISE), Bld. 38, Lenina Ave, Magnitogorsk, Russia, 455000. E-mail: inna_barankova@mail.ru;

MIKHAILOVA Uliana, NMSTU, Ph.D.,Associate Professor of CSISE Department, Bld. 38, Lenina Ave, Magnitogorsk, Russia, 455000, E-mail: ylianapost@gmail.com;

LUKIANOV Georgy, NMSTU, Teaching Assistant of CSISE Department, Bld. 38, Lenina Ave, Magnitogorsk, Russia, 455000, E-mail: decorsi@mail.ru.