ИССЛЕДОВАНИЕ ТЕКУЩИХ ДОКУМЕНТОВ ВЫЯВЛЯЕТ СЕКРЕТЫ

По мнению американского исследователя, многие документы, публикуемые в стиле online (без специальной обработки), могут неявно содержать конфиденциальную корпоративную или личную информацию.

Саймон Баерс, из исследовательской лаборатории компании AT&T, смог извлечь скрытую информацию из многих тысяч документов Microsoft Word, используя широко доступные программные средства и стандартную технику программирования.

Профессиональные редактирующие программы часто сохраняют в файле ту информацию, которую конечный пользователь не видит. Действительно, сохранение удаленных фрагментов текста делает редактирование более эффективным. Однако, по мнению Баерса, это может подвергнуть ничего не подозревающего пользователя значительному риску. В своем докладе Баерс утверждает, что злоумышленник, анализируя электронные документы, мог бы овладеть информацией, которую можно использовать в промышленном шпионаже или при мошенничестве. Вполне реально, что кто-то включит свой номер социального страхования в копию резюме, посланного возможному работодателю, но удалит его из окончательной версии, передаваемой *online,* надеясь таким образом сохранить конфиденциальную информацию.

Используя обычную поисковую программу и случайный набор ключевых слов, Баерс смог обнаружить более 100 000 документов Word, содержащих деловые документы и индивидуальные резюме. Он выбрал документы Word в силу их распространенности, но подчеркнул, что документы других форматов могут содержать аналогичную скрытую информацию. Например, в 2002 году газета Washington Post опубликовала версию письма, посланного «вашингтонским снайпером» в pdf-формате. Номера и телефоны были удалены из видимого текста, но все еще содержались в файле. Однако новое исследование, проведенное Баерсом, показало, насколько распространена эта проблема.

Баерс использовал программы открытого доступа «antiword» и «catdoc» для преобразования word-файлов в простые текстовые файлы. Затем он написал простой сценарий для выявления того текста, который не виден в формате Word'а. Баерс обнаружил большое количество удаленного текста и массу конфиденциальной информации (заголовки писем, имена людей, адреса, отрывки из обсуждаемых документов). Neil Laver, руководитель группы маркетинга продуктов Microsoft Office в Великобритании, сказал, что компания разрабатывет способы, позволяющие потребителям закрыть от постороннего доступа конфиденциальную информацию, оставленную в файле в скрытом виде. Он сказал, что скрытая информация чрезвычайно полезна для улучшения функциональности программ. Однако, по его словам, если какая-то часть данных является конфиденциальной, то должны существовать способы предотвращения

их несанкционированного распространения. Следующая версия Office 2003 будет содержать средства, позволяющие пользователям удалять личную информацию из документа. Она также будет включать «управление правом на информацию», которое даст возможность автору определять, кто имеет право читать или пересылать документ.

Другие программы могут быть использованы для удаления скрытого текста из документов, но пока что, возможно, самый лучший способ – это перед публикацией ковертировать документ в простой ASCII текст.

Статья Баерса представлена для публикации в IEEE journal Security and Privacy.

Will Knight
NewScientist.com news service
(www. NewScientist.com)

От редакции:

Мы запустили программу «antiword», чтобы проверить сообщение автора.

Однако мы не обнаружили описанного эффекта. Возможно, что было что-то упущено при запуске программы, поэтому мы решили разместить ее на диске в приложении к журналу, чтобы читатели также смогли попробовать, и, возможно, они будут более удачливы.

Этой же теме была посвящена заметка М. Хмелева в газете «Известия» за 12.09.03 г. под названием «Агент на рабочем столе» с подзаголовком «Word не гарантирует безопасность приватной информации». Статья была воспроизведена во многих электронных СМИ. В ней автор не только описывает, ссылаясь на того же С. Баерса, нежелательные последствия, представляемых редактором Word возможностей увидеть предыдущие версии текста, но и дает рекомендации, как исключить утечку конфиденциальной информации.

В заметке, в частности, автор пишет: «Основную опасность для пользователя представляют две функции программы – «отслеживание исправлений» и режим «быстрого сохранения текста». В режиме слежки за изменениями программа автоматически сохраняет все исправления и комментарии, сделанные в процессе подготовки документа. Все произведенные вами изменения текста, сохраненного в формате .doc, в том числе и не вошедшие в окончательную редакцию, вполне доступны даже без применения каких-либо дополнительных программных средств.

«В моей практике был неприятный случай – в процессе подготовки договора с одним из наших американских контрагентов мой коллега сделал на полях электронного документа несколько нелицеприятных комментариев о нашем партнере, – рассказывает Игорь Ашманов, гендиректор компании «Ашманов и Партнеры», один из разработчиков «проверки русского правописания» в Word. В итоге, получив от нас файл .doc с проектом договора, наш заокеанский партнер узнал о себе много нового – просто у него была установлена другая версия программы с другими настройками».

Если пользователем включен режим «быстрого сохранения», Word сохраняет не весь текст сразу, а только те изменения, которые были внесены с момента последне-

...получив от нас файл .doc с проектом договора, наш заокеанский партнер узнал о себе много нового.

го сохранения. Однако «быстрое» в данном случае не значит безопасное. Для поддержания этой функции внутри файла .doc создается своеобразный «отстойник», где и сохраняются все последние изменения документа. Недоброжелатель, немного знакомый с технической кухней программы, сделав глубокий анализ документа, с легкостью получит доступ ко всем предварительным редакциям текста. Известен случай, когда один из бизнесменов, получив от своего партнера проект контракта по электронной почте, смог получить доступ к информации по четырем аналогичным контрактам своего визави за последние два года.

О том, что у Word'а большие проблемы с сохранением приватности, известно уже лет десять, и исключить утечку подобных данных возможно, – говорит Игорь Ашманов. – Давно уже написан и доступен бесплатный софт, позволяющий подчистить в документе все хвосты – следы изменений и персональные данные. Есть и более доступный способ – перед пересылкой любой мало-мальски важный документ конвертировать в RTF-формат, который помимо того, что не сохраняет никаких скрытых данных, абсолютно гарантирует от вирусов».