Парламентское собрание Союза Беларуси и России

Постоянный Комитет Союзного государства

Оперативно-аналитический центр
при Президенте Республики Беларусь

Государственное предприятие «НИИ ТЗИ»

Полоцкий государственный университет

КОМПЛЕКСНАЯ ЗАЩИТА ИНФОРМАЦИИ

Материалы XXII научно-практической конференции

(Полоцк, 16–19 мая 2017 г.)

Новополоцк
2017

УДК 004(470+476)(061.3)
ББК 32.81(4Беи+2)
К63

К63 **Комплексная защита информации :** материалы XXII науч.-практ. конф., Полоцк, 16–19 мая 2017 г. / Полоц. гос. ун-т ; отв. за вып. С. Н. Касанин. – Новополоцк : Полоц. гос. ун-т, 2017. – 282 с.
ISBN 978-985-531-564-4.

В сборнике представлены доклады ученых, специалистов, представителей ггосударственных органов и практических работников в области обеспечения информационной безопасности Союзного государства по широкому спектру научных напрвлений.

Адресуется исследователям, практическим работникам и широкому кругу читателей.

Тексты тезисов докладов, вошедших в настоящий сборник, представлены в авторской редакции.

УДК 004(470+476)(061.3)
ББК 32.81(4Беи+2)

ISBN 978-985-531-564-4

ва. Для повышения уровня безопасности личного информационного пространства эксперта для доступа к информационным системам, обрабатывающим информацию ограниченного распространения, интересным представляется защищенное программно-аппаратное устройство на основе Гарвардской архитектуры, предложенное В. Конявским и В. Степановым, описанное в патенте №118773.27.07.12. Основное отличие новой Гарвардской архитектуры от архитектуры фон Неймана заключается в том, что в компьютере с использованием новой Гарвардской архитектуры процессор может читать инструкции и выполнять доступ к памяти данных одновременно, без использования кэш-памяти. Таким образом, компьютер с новой Гарвардской архитектурой при определенной сложности схемы быстрее и безопаснее, чем компьютер с архитектурой фон Неймана, поскольку шины инструкций и данных расположены на разных, не связанных между собой физически, каналах. Исходя из физического разделения шин команд и данных, разрядности этих шин (следовательно, и адресные пространства) могут иметь различные значения и физически не могут пересекаться друг с другом [3,4].

Таким образом, объединяя удачные технические и программные решения, становится возможной безопасная работа с конфиденциальной информацией без потери мобильности и удобства современных вычислительных сред. Благодаря данному подходу к проектированию личного информационного пространства, человек получает компромиссный вариант управления персональной информацией в условиях современного кризиса личной информационной безопасности.

Список литературы

1. Таненбаум Э. Распределенные системы. Принципы и парадигмы. – СПб.: Питер, 2003. – 877 с.

2. Фрэнкс Б. Укрощение Больших данных. М.: Манн, Иванов и Фербер, 2014. 352 с.

3. Конявский В. А., Кузнецов Д., Волчков А., Афанасьев А. Специализированные компьютеры – панацея от хакеров? // Аналитический банковский журнал. – No12 (224) декабрь 2014. – С. 66–69.

4. Конявский В. А., Степанов В. Б. Компьютер типа «тонкий клиент» с аппаратной защитой данных. Патент на полезную модель No 118773.27.07.12, бюл. No21

КЛАССИФИКАЦИЯ И РАЗМЕТКА ЭЛЕМЕНТОВ ТЕКСТА НА УРОВНЕ ПРЕДЛОЖЕНИЯ ДЛЯ ВЫЯВЛЕНИЯ УЯЗВИМОСТЕЙ НУЛЕВОГО ДНЯ

Е.Д. МАТЯШ

Московский технологический университет

Введение

Сейчас стало популярным использование таких систем, которые так или иначе имеют дело с понятием BigData – в данном случае имеется ввиду разметка текста в плане его разделения на различные элементы разных типов: слова, предложения, абзацы и т.д. Это разделение необходимо для поддержания работы многих систем, которые напрямую связаны с текстовой информацией. Текстовые редакторы, поисковики и любые другие информационные системы работают непосредственно за счет специальных алгоритмов обработки исходной текстовой информации. Все эти алгоритмы,

в первую очередь, осуществляют морфологическую разметку текста, т.е. каждому отдельному слову присваивается специальный тег, в котором хранится вся его морфологическая характеристика, включая часть речи.

Разметка текста представляет собой разделение текста на более мелкие составляющие: сначала на абзацы, затем на предложения, после – на слова, вплоть до определения части речи и значения каждого слова. Такая процедура необходима не только для разработки и усовершенствования всяческих пользовательских приложений, но и для обеспечения безопасности различных систем.

Число создаваемых новых технических продуктов с каждым годом неуклонно растет. Это обусловлено тем, что наука не стоит на месте - улучшая и дополняя имеющиеся изобретения, создает совершенно уникальные творения. Однако какой бы не была система защищенной и какие бы меры безопасности не принимались разработчиками - все равно найдутся такие люди, которые смогут обнаружить уязвимости, не оставив время девелоперам на hotfix ("горячее исправление").

Принципы построения методов разбора текста при поиске информации об уязвимостях нулевого дня.

Уязвимости нулевого дня – не устраненные уязвимости, а также вредоносные программы, против которых ещё не разработаны защитные механизмы. Поэтому для поиска таких уязвимостей можно использовать уже опубликованную информацию.

При поиске информации об уязвимостях нулевого дня необходимо решить ряд задач обработки естественного языка:

1. Поиск фрагментов текста – разделение материала на различные элементы разных типов: слова, предложения, абзацы и. т. д.

2. Поиск предложений (Sentence Boundary Disambiguation, SBD) – определение границ предложения.

3. Поиск именованных объектов (Named entity recognition, NER) – механизм поиска названий компонентов, программного обеспечения, процессов, конкретных уязвимостей, или любых других именованных сущностей.

4. Определение частей речи (Parts of speech, POS) – классификация элементов текста на уровне предложения. Предложение может быть разделено на отдельные слова и словосочетания по таким категориям, как существительные, глаголы, наречия, предлоги и т.д.

5. Классификация текстов и документов – цель данной классификации в присвоении меток фрагментам, найденным в текстах и документах.

6. Выделение взаимоотношений – выявление связей между словами или словосочетаниями для построения семантического дерева.

При анализе открытых источников могут быть использованы словари об уязвимостях в постах социальных сетей и блогов:

СПО? уязвимость;
СПО? взлом;
СПО? взломан;
СПО? утечка;
СПО? злоумышленники;
СПО? перехват;
СПО? читать чужую;
СПО? хакеры;
СПО? хакеры прочитать;
СПО? хакеры взломали;
СПО? хакеры аккаунты;

СПО? хакеры сотни;
СПО? хакеры тысячи;
СПО? хакеры переписку;
СПО? красть;
СПО? украсть;
СПО? скомпрометированы.

К базовым задачам обработки текста относятся: токенизация (разбиение текста на осмысленные элементы - слова, фразы, символы), стемминг (процесс нахождения основы слова, которая не обязательно совпадает с корнем слова), лемматизация (приведение слова к словарной форме), определение границ предложений и фильтрация стоп-слов.

Разбор текста осуществляется при помощи специализированных программ, которые выполняют свою задачу посредством трех инструментариев: заранее подобранный словарь слов и словосочетаний (для более углубленного анализа), правила, машинное обучение (во избежание противоречивых смысловых накладок). Далее эти принципы будут детально описаны.

Первый принцип – это словарь. Если слова в словаре нет, то морфологический анализ не может быть выполнен, и, как следствие, не могут быть определены грамматические характеристики слова. Большинство разработчиков использует в качестве базы словарь Зализняка, причем работы, сопоставимой с его, еще никто не делал, хотя прошло уже довольно много времени с его создания.

Второй принцип – специальный набор правил морфологического разбора слов, включающий в себя словообразующие морфемы.

Третий принцип – машинное обучение, которое является классом методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решения множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме.

Заключение

На сегодняшний день не существует русскоязычного процессора, в котором все известные проблемы были бы решены до конца.

Была рассмотрена возможность поиска уязвимостей нулевого дня с применением информации в Интернет. Для анализа информации на Интернет-страницах, написанных на естественном языке, использовались подходы токенизации и определения вероятности принадлежности к категории описания уязвимости.

Дальнейшие исследования должны быть направлены на подбор обучающей выборки и разработку программного обеспечения на различные Интернет-ресурсы, содержащие информацию об уязвимостях нулевого дня.

Список литературы

1. Зализняк А.А. Грамматический словарь русского языка. Словоизменение. – М., «Русский язык», 1977, 880 с.

2. Луканин А.В. Автоматическая обработка естественного языка /; М-во образования и науки Российской Федерации, Южно-Уральский гос. ун-т, Каф. Общая лингвистика». – Челябинск: Изд. центр ЮУрГУ, 2011, 70 с.

3. Леонтьева Н.Н. Автоматическое понимание текста: системы, модели, ресурсы: учебное пособие – М.: Издательский центр «Академия», 2006, 304 с.

4. Севбо И.П. Структура связного текста и автоматизация реферирования / М.: Наука, 1969, 135с.

5. Jurafsky D. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition / D. Jurafsky, J.H. Martin. – New Jersey: Prentice Hall, 2000, 934p.

О ФОРМИРОВАНИИ ЦЕНТРАЛИЗОВАННОГО АУДИТА СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В СОСТАВЕ ПОДСИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИОННЫХ СИСТЕМ РЕСПУБЛИКИ БЕЛАРУСЬ

А.О. МОЛЧАН, О.Э. СЯЗАНЦЕВ, Н.В. ЖУРАВСКИЙ

Научно-производственное республиканское унитарное предприятие «Научно-исследовательский институт технической защиты информации»

В силу быстро растущей информатизации общества количество хранимой, обрабатываемой и передаваемой в электронном виде информации растет в геометрической прогрессии. Как следствие, все в большем масштабе поднимается проблема общедоступности информации, а также взаимодействия с информацией, распространение и (или) представление которой ограничено. Возникает острая необходимость в контроле и ограничении доступа к обрабатываемой информации, а также в своевременной информированности ответственных за информационную безопасность (далее – ИБ) лиц об инцидентах ИБ. Помимо общедоступности информации уязвимость информационных систем (ИС) возрастает также за счет повышения сложности элементов ИС, появления новых технологий передачи и хранения данных, увеличения объема программного обеспечения (ПО).

Для мониторинга ИС на предмет корректного функционирования всех ее процессов, активов и ресурсов, а также сохранности информации, доступ к которой ограничен, в составе системы защиты информации (СЗИ) ИС организуется подсистема аудита событий ИБ.

Системы аудита событий ИБ используются для мониторинга ИС на предмет корректного функционирования и сохранности информации, доступ к которой ограничен. На данный момент существует множество средств аудита событий ИБ, как встроенных в общесистемное ПО, так и специально разработанных программных и программно-аппаратные комплексов компаниями и корпорациями разных стран. Однако в государственных ИС согласно законодательству Республики Беларусь в области защиты информации (ЗИ) допустимо использование только тех средств ЗИ, которые прошли сертификацию по требованиям безопасности информации и получили сертификат соответствия либо экспертное заключение на соответствие требованиям технических нормативных правовых актов в области технического нормирования и стандартизации.

Многие средства ЗИ ведут журналы аудита автономно, и интеграция их производится вручную ответственными за ИБ лицами. Главная проблема формирования централизованного аудита ИБ в составе СЗИ ИС на данный момент является дороговизна иностранных продуктов (в частности, систем SIEM) и ограниченное представление систем аудита ИБ белорусского производства на рынке средств ЗИ, которые также позволяли бы снизить угрозы ИБ.