Чечулин А. А., Зозуля Ю. В., Котенко И.В., Тишков А. В., Шоров А. В.

МЕТОДЫ ЗАЩИТЫ ОТ ВРЕДОНОСНЫХ WEB-САЙТОВ НА ОСНОВЕ ОЦЕНОК РЕПУТАЦИИ

Санкт-Петербург, ООО "Центр специальной системотехники", СПИИРАН
chechulin@comsec.spb.ru, yuzozulya@gmail.com,
ivkote@comsec.spb.ru, avt@iias.spb.su, ashorov@comsec.spb.ru

Одним из классов методов, которые могут использоваться для защиты от вредоносных Web-сайтов, является так называемые репутационные методы. Эти методы базируются на понятии репутации, отражающей некоторую меру вредоносности сайта, определяемую на основе множества источников.

В докладе проводится анализ существующих методов оценки репутации и предлагается подход к оценке репутации Web-сайтов, формируемый на основе многофакторного анализа его характеристик, получаемых из различных источников.

Можно выделить несколько классов методов вычисления репутации. Наиболее известными и используемыми являются методы, основанные на подсчете средних значений, потоковые модели и системы байесовского типа.

Например, стандартным примером метода, базирующегося на подсчете средних значений, является Интернет-аукцион eBay. Другие примеры - Amazon, Epinions.

Потоковые модели основаны на графах, отображающих ссылки веб-страниц друг на друга или связи между другими объектами. Наиболее распространены итеративные алгоритмы распространения рейтингов по графу смежности, представляющему сеть. Использование теорем линейной алгебры и теории марковских цепей о сходимости итеративного пересчета рейтингов при случайном блуждании по графу смежности к собственному вектору соответствующей матрицы, позволяет получить приблизительные оценки рейтингов, сколь угодно близкие к конечным предельным значениям.

Наиболее известным среди потоковых алгоритмов является алгоритм PageRank [1]. Он используется, в частности, Google для оценки рейтинга веб-сайтов. Узлам сети, представляющей собой структуру переходов между страницами, приписываются вероятности перехода пользователя от одной страницы к другой. Используя механизм случайных блужданий, вычисляются рейтинги узлов, как предельные значения вероятности нахождения в конкретном узле.

В персонализованном PageRank [2] предлагается подход, позволяющий добавить предпочтения пользователей к общему алгоритму PageRank. Алгоритм Hyperlink-Induced Topic Search (HITS) [3] так же как и PageRank, строит ориентированный граф по связям между страницами. Однако, одним из базовых отличий является работа на ограниченном наборе страниц, базирующаяся на результатах поискового запроса. Алгоритм Stochastic Approach for Link-Structure Analysis (SALSA) [4] является обобщением алгоритма HITS. Модификация заключается в том, что строятся две матрицы смежности, отдельные для авторитетностей и хабов. Алгоритм TrustRank [5] представляет способ экспертной оценки нескольких узлов как (не)источников спама, результаты которых распространяются по сети. Алгоритм EigenTrust [6] рассматривает peer-to-peer (P2P) сети, в которых узлы одновременно являются поставщиками и приемниками файлов (в принципе, можно рассмотреть любые транзакции между эквивалентными узлами и понятие успешности транзакций).

Известны также потоковые методы для более сложных структур.

К другим способам оценки репутации можно отнести дискретные модели доверия, нечеткие модели, модели мнений и др.

В широко используемых в настоящее время системах защиты от вредоносных Web-сайтов, прежде всего родительского контроля (Cyber Patrol, Net Nanny и др.), понятие

вредоносности определяется путем отнесения сайта к одной или нескольким нежелательным категориям (например, «накротики», «насилие», «порнография» и др.).

Для вредоносных систем характерным является наличие базы гиперссылок (URL), которые используются для причинения вреда [7-9]. Гиперссылки применяются источниками спама, вирусами, распространяющимися через Web страницы, fishing-сайтами, шпионским ПО и т.д. Построение и анализ системы рейтингов для этих URL-адресов позволит заблокировать опасные Web-сайты на основе их репутации и тем самым избежать большого количества угроз.

При анализе репутация строится на основе многих источников данных, которые можно разделить на две категории: автоматические и пользовательские. Среди автоматических источников также можно выделить две подкатегории: анализаторы (содержимого, размещения, истории и т.д.), использующие собственные алгоритмы анализа, и общедоступные внешние источники, предоставляющие “черно-белые” списки и/или результаты оценки сайтов.

Основной идеей предлагаемого подхода является использование в качестве исходных данных не только результатов локального анализа содержимого сайта, но и контекстной информации, включающей в себя историю сайта, оценки содержимого сайта из внешних хранилищ данных и т.п. Процесс обновления всех баз, содержащих данные подобного рода, на хосте конечного пользователя является весьма ресурсоемким. Поэтому практически всеми производителями систем нового поколения применяется концепция «умного облака» (Smart Cloud) [7], которое запрашивается при необходимости посредством Internet – соединения.

Для создания эффективной системы репутации, основанной на собственных анализаторах, важно определить данные, прямо или косвенно выражающие злонамеренность содержимого, а также отражающие адекватность имеющейся информации о сайте. Такими данными могут быть данные об истории сайта (возраст, страна, в которой зарегистрирован сайт, организация, предоставляющая хостинг сайту, история серверов, на которых размещался сайт), анализ связей между сайтами (анализ репутаций сайтов, на которые ссылается исследуемый сайт, анализ репутаций сайтов, ссылающихся на исследуемый сайт), анализ содержимого сайта на наличие объектов заданных категорий (анализ текстовой информации на наличие ключевых слов и фраз, анализ изображений представленных на сайте, анализ новых процессов в системе, появившихся после посещения сайта, антивирусный анализ сайта).

Достоинства разработки собственных анализаторов: высокая производительность, не требует дополнительных расходов (кроме возможных расходов на оборудование), возможность оценки большинства запрашиваемых сайтов. К недостаткам можно отнести возможность обмана системы (создание сайта таким образом, чтобы автоматическая система давала некорректную оценку), ложные срабатывания (отнесение в запрещенные категории хороших сайтов), требуется мощное оборудование при большом количестве запросов к сайтам, устаревание оценки.

Данные, которые необходимо собирать и постоянно обновлять с внешних источников, могут представляться регулярно обновляемыми списками сайтов по категориям от сторонних источников (в том числе платные каталоги), регулярно обновляемыми “черно-белыми” списками сайтов по Parent Control, Fishing и т.д., внешними системами оценки вредоносных сайтов (Google safebrowsing и т.д.).

Достоинства получения информации от внешних источников: высокая точность (особенно коммерческих списков), высокая производительность. Однако имеется и ряд недостатков, прежде всего разные критерии оценки и противоречивость списков разных источников, а также устаревание оценки, платный доступ к хорошим и регулярно

обновляемым спискам, желательна проверка качества списков, неполнота списков (т.е. список не может охватывать все существующие сайты).

Оценки сайта пользователями (группами экспертов по областям знаний и сообществом пользователей с учетом степеней доверия к ним) важны, но работа по составлению рейтингов вручную весьма трудоемка. Поэтому пользовательская оценка может быть использована в качестве важного дополнения к автоматической. Прежде всего, следует использовать экспертов-специалистов, составляющих небольшую группу пользователей, которым система оценки репутации доверяет.

Очевидные достоинства оценки пользователей - это высокая точность оценки и решение спорных вопросов. Экспертные оценки имеют наибольший вес (в случае использования весового алгоритма объединения оценок) при вычислении репутации сайта; на их основе можно создавать свои списки и правила разных городов и стран, а также обеспечить быстрое внесение в список недавно появившихся сайтов.

Среди недостатков отметим низкую производительность, небольшое число экспертов, разные критерии оценки у разных пользователей, устаревание оценки (необходимо учитывать динамику изменения содержимого ресурсов), возможность умышленной фальсификации отзывов от обычных пользователей.

Вся информация от разных источников должна комбинироваться, и вычисление репутации ресурса должно осуществляться по каждой категории. При этом оценки пользователей могут комбинироваться с результатами автоматического анализа и с существующими оценками при помощи весовых коэффициентов, причем при расчетах может учитываться коэффициент устаревания оценки ресурса источниками. Результатом работы системы оценки являются оценки сайта по набору заранее заданных категорий. Причем оценка идет не по принадлежности/не принадлежности категории, а с более детальной градацией (например, с использованием оценки от 0 до 100).

Данная работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (РФФИ) (проект №07-01-00547) и программы фундаментальных исследований ОНИТ РАН.

Литература

1. Brin S. Lawrence Page: The Anatomy of a Large-Scale Hypertextual Web Search Engine // Computer Networks, Vol. 30, No.1-7, 1998.
2. Jeh G., Widom J. Scaling personalized web search // Proceedings of the Twelfth International World Wide Web Conference. 2002.
3. Kleinberg J. M. Authoritative sources in a hyperlinked environment // Journal of the ACM, Vol. 46, No. 5, 1999.
4. Lempel R., Moran S. SALSA: the stochastic approach for link-structure analysis // ACM Transactions on Information Systems, Vol. 19, No. 2, 2001.
5. Gyöngyi Z., Garcia-Molina H., Pedersen J. Combating web spam with trustrank // Proceedings of the Thirtieth international conference on Very large data bases, 2004.
6. Kamvar S. D., Schlosser M. T., Garcia-Molina H. The eigentrust algorithm for reputation management in p2p networks // Proceedings of the 12th International WWW Conference, 2003.
7. Securecloud. Trend Micro. https://securecloud.com/
8. Web Reputation Solutions from SECURE COMPUTING. http://www.securecomputing.com/gateway/web_reputation.cfm
9. Iron Port Web Reputation: Protect and Defend Against URL-based Threats. http://www.ironport.com/pdf/ironport_web_reputation_whitepaper.pdf