УДК 004.056.57

Н.Г. Булахов

Методы обнаружения компьютерных вирусов и сетевых червей

Представлена и описана статистическая модель цифровой информационной сети, применимая для обнаружения факта распространения вредоносного сетевого программного обеспечения.

По сведениям аналитиков американской компании PC Tools, Россия сейчас является лидером в распространении компьютерных вирусов, вредоносного и шпионского программного обеспечения [1]. Поэтому актуальная задача сегодня — анализ методов обнаружения компьютерных вирусов и червей с целью разработки новых альтернатив. Рассмотрим три основных класса методов: сигнатурный, статистический и эвристический. В отдельный класс следует выделить новый метод, основанный на анализе характеристик (а не содержимого) передаваемого по сети трафика, который практически одновременно предложен в работе [2] и автором [3].

Сигнатурные методы анализа [4] описывают каждую атаку индивидуальной моделью, или сигнатурой. Ею могут служить строка символов, семантическое выражение, формальная математическая модель и т. д. [5].

Продукционные / экспертные системы обнаружения вторжения кодируют данные об атаках и правила импликации «если ... то», а также подтверждают их, обращаясь к контрольным записям событий.

Обнаружение вторжений, основанное на модели, — один из вариантов, объединяющий модели вторжения и доказательств, поддерживающих вывод о вторжении. В системе обнаружения вторжений поддерживается база данных сценариев атак.

Анализ перехода системы из состояния в состояние осуществлён в системах STAT и USTAT под ОС UNIX. В них обнаружения вторжений атаки представляются как последовательность переходов контролируемой системы из состояния в состояние.

Изменение состояний и сети Петри для обнаружения вторжений развивают метод анализа изменений состояний. Хотя данный метод позволяет получить более точные результаты в области обнаружения вторжений, он требует значительных вычислений.

Статистические методы анализа предназначены для выявления безопасности поведения программ и систем обнаружения нарушителя [5].

Операционная модель основывается на том, что каждое новое наблюдение переменной должно укладываться в некоторых границах. Если этого не происходит, то имеет место отклонение.

Модель среднего значения μ ***и среднеквадратичного отклонения*** σ базируется на том, что всё знание о предыдущих наблюдениях некоторой величины есть величины μ и σ. Тогда новое наблюдение является аномальным, если оно не укладывается в границах доверительного интервала $\mu + d^*\sigma$, где d^* — априори устанавливаемая величина.

Многовариационная модель аналогична модели среднего значения и среднеквадратичного отклонения, но учитывает корреляцию между двумя или большим количеством метрик (использование центрального процессорного устройства и количество операций ввода-вывода, количество выполненных процедур входа в систему и время сессии).

Модель марковского процесса [5, 7] применима только к счётчикам событий, если рассматривать каждый тип событий как переменную состояния и использовать матрицу переходов для характеристики частот переходов между состояниями. Результат наблюдения является аномальным, когда вероятность перехода, определённая предыдущим состоянием и матрицей перехода, очень мала.

Модель временных серий использует временные периоды вместе со счётчиками событий и измерениями ресурса, учитывая как значения наблюдений, так и временные интервалы между ними. Новое наблюдение считается аномальным, если вероятность его появления (с учётом длительности наблюдения) низка.

Метод информационного анализа [5, 7] выделяет особенности больших наборов данных. Здесь применялась система RIPPER (Repeated Incremental Pruning to Produce Error

Reduction), построенная на основании обучающих правил и используемая для решения задач классификации.

Метод конечных автоматов состоит в разработке конечного автомата для распознавания «языка» трассы программы. Для этого существует много методик, основанных на использовании как детерминированных, так и вероятностных автоматов.

Эвристические методы. Программу, которая анализирует код проверяемого объекта и по косвенным признакам определяет, является ли объект вредоносным, называют эвристический анализатор (эвристик) [8].

Метод «песочниц» таков: подозрительный файл помещают в область, изолированную от остальной системы, проверяя его поведение. При этом каждый исполняемый файл запускается в виртуальной среде, которую стандартный программный интерфейс приложений Windows (Application Programming Interface) организует в так называемой тюрьме (jail) [6].

Метод блокировки поведения комбинирует оба упомянутых подхода. Благодаря иерархической обработке трафика данных, высокая потребность «песочницы» в ресурсах сводится к минимуму и уменьшается вероятность ложных срабатываний эвристических методов [6].

Метод мониторинга характеристик передачи сетевого трафика. Перспективен способ обнаружения быстро распространяющихся вирусов и червей, основанный на постоянном мониторинге сетевого трафика. Подобную технологию реализуют исследователи из Пенсильванского университета [2]. При отслеживании параметров передачи сетевого трафика система анализирует количество пакетов данных, пересылаемых между различными сетями, и в случае обнаружения аномального всплеска активности подаёт сигнал тревоги. Это позволяет идентифицировать вирусную эпидемию в течение долей секунды после её начала.

Затронутые выше методы систематизированы в таблице, где символом «+» отмечены методы с наилучшими характеристиками по данному параметру, символом «–» — методы с наихудшими, «*» — со средними. Сравнение взято из литературы.

Классификация методов анализа по параметрам, критичным к обнаружению атак вредоносных программ

Классы методов анализа	Методы анализа	Параметры									
		Ресурсоёмкость	Время выявления	Эффективность на ранней стадии	Эффективность на поздней стадии	Ложные срабатывания	Универсальность	Наличие обновляемых баз	Простота критериев оценки	Самостоятельность метода	Необходимость обучения системы
Сигнатурный	Продукционные / экспертные системы обнаружения атак	*	+	–	+	+	–	–	–	+	+
	Обнаружение вторжений, основанное на модели	*	+	–	+	+	–	–	–	–	+
	Анализ перехода системы из состояния в состояние	+	+	–	+	+	–	–	–	+	+
	Изменение состояний и сети Петри	–	–	–	+	+	–	–	+	+	+
Статистический	Статистический анализ последовательности системных вызовов	+	+	+	–	–	+	+	–	+	–
	Конечные автоматы	–	–	+	–	–	+	+	–	+	–
Эвристический	Анализ поведения системы	+	+	*	+	*	–	+	–	+	+
Мониторинг активности	Анализ интенсивности передачи сетевых пакетов	+	+	+	+	–	+	+	+	+	+

Из таблицы правомерно заключить, что предлагаемый нами метод мониторинга отличается высокой конкурентоспособностью.

Для анализа данных мониторинга сетевого трафика необходимо исследовать поведение вредоносных программ в цифровых информационных сетях (ЦИС). Для описания функционирования ЦИС традиционно применяют эпидемиологические модели [9, 10]. Но они позволяют лишь наблюдать динамику роста числа заражённых машин и не отображают важных сетевых параметров: способность сети передавать пользовательскую и служебную информацию и т.д. Поэтому актуальна разработка модели, учитывающей пиковые нагрузки в ЦИС при передаче пакетов информации и использующей статистическое описание этих процессов.

Мы предлагаем модель, основу которой составляет устройство пересылки пакетных данных, имеющее N входов и N выходов (рис. 1). Входы и выходы не равноправны между собой, что позволяет отразить наличие в реальной сети магистральных каналов и присоединение оконечных станций. Для учёта различия между ними каждому каналу присваивается свой весовой коэффициент $W = WO+WV$, который показывает, насколько вероятнее появление пакета на данном входе относительно всех остальных. Вес WO соответствует нагрузке канала за счёт полезных пакетов, вес WV — за счёт саморазмножающихся пакетов. Входящие пакеты имеют адрес назначения, указывающий выходной порт. Каждый выходной порт рассчитан на очередь длиной M пакетов. Индекс max обозначает максимальный вес.

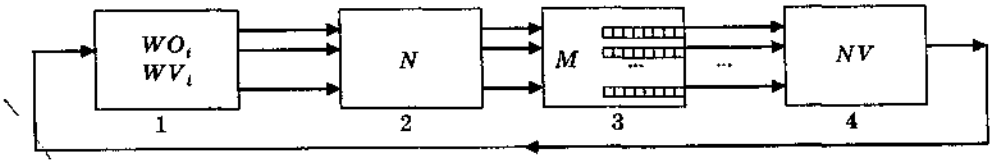


Рис. 1. Модель передачи данных в ЦИС с блоками распределения пакетов на входе (1) и перераспределения пакетов по выходам (2), очередей (3), обработки информации о выходных пакетах (4)

В заданный такт времени условие того, что на входе n есть пакет, определяется формулой (где символ rnd — случайная величина из [0; 1])

$$f(n) = \begin{cases} 1, \text{rnd} \le \dfrac{WO_i + WV_i}{WO_{\max} + WV_{\max}}, \\ 0, \text{rnd} > \dfrac{WO_i + WV_i}{WO_{\max} + WV_{\max}}. \end{cases} \quad (1)$$

Если на входе n имеется пакет, то условие того, что он саморазмножающийся либо полезный, есть

$$f_1(n) = \begin{cases} 1, \text{rnd} \le \dfrac{WV_i}{WO_i + WV_i}, \\ 2, \text{rnd} > \dfrac{WV_i}{WO_i + WV_i}. \end{cases} \quad (2)$$

В зависимости от порта назначения k каждый пакет помещается в выходную очередь согласно условию

$$\sum_{i=1}^{k} (WO_i + WV_i) \le \text{rnd}\left(\sum_{i=1}^{N} (WO_i + WV_i) \right) < \sum_{i=1}^{k+1} (WO_i + WV_i). \quad (3)$$

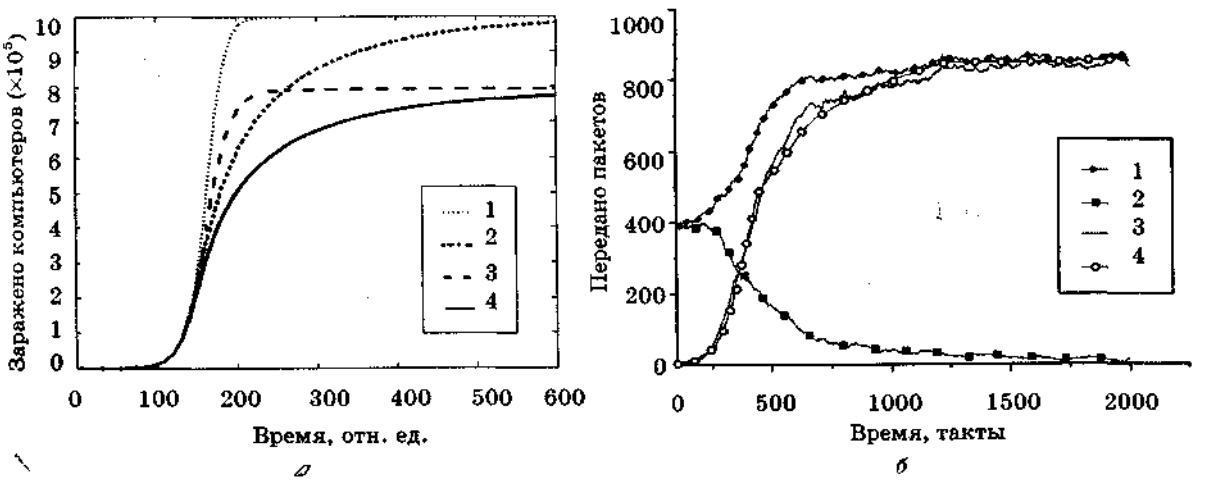
А если число пакетов в очереди достигает M, то $(M+1)$-й отбрасывается.

Между входом и выходом устройства есть положительная обратная связь. Тогда при увеличении числа заражённых пакетов на выходе увеличивается вероятность появления заражённого пакета на входе. Согласно модели подсчитывается число пакетов каждого типа (NO_t и NV_t) на выходе. Коррекция весового коэффициента $WV_{i,t}$, соответствующего входу i, в такт времени t выполняется согласно соотношению

$$WV_{i,t} = WV_{i,t-1} \frac{NV_t + 1}{NV_{t-1} + 1}. \quad (4)$$

Постулируем линейную связь между количеством заражённых червями компьютеров и порождаемым ими трафиком. Сравнивая результаты моделирования на базе эпидемиологической модели и её модификаций [9], а также статистические данные об атаке сетевого червя *Code Red* [9] с результатами моделирования на основе предложенной модели [11] (рис. 2), можно сделать вывод об её адекватности.

Масштаб соотнесения 200 тактов в час. Предполагается, что один заражённый компьютер генерирует 440 вредоносных пакетов в такт.



а *б*

Рис. 2. Распространение червя: *а* — симуляция с помощью простой модели (1), с учётом замедления скорости заражения (2), с учётом предпринимаемых людьми контрмер (3), *двухфакторной модели* [9] (4); *б* — симуляция работы ЦИС на основе предложенной модели: динамика общего количества переданных пакетов (1), пакетов с полезными сообщениями (2), *вредоносных пакетов (3) и статистические данные о CodeRed [9] (4)*

Выводы. Для выявления потенциально опасной активности в ЦИС индикаторами служат: значения интенсивности пересылки отдельных пакетов, небольших очередей пакетов и попыток установить соединение отдельными хостами; распределение IP- и MAC-адресов источника и назначения в передаваемых пакетах; их размеры и тип. Но для повышения надёжности детектирования вредоносной активности и сведения риска ложного срабатывания к минимуму в данном случае часто требуется некоторая «калибровка», составление сигнатур распространения известных червей. А это делает сети уязвимыми для новых (не внесённых в базу) червей.

Предлагается исключить параметры, варьирующиеся для конкретных реализаций червя, и добавить характеристики пересылки информационных пакетов внутри маршрутизаторов и коммутаторов: количество отбрасываемых пакетов в единицу времени, заполненность буферов пересылки, нагрузку оборудования. Тем самым можно увеличить количество идентификаторов сетевой атаки. Предложенная статистическая модель ЦИС, учитывающая возможность пиковых нагрузок, адекватна. Об этом говорят представленные результаты верификации модели.

Литература

1. Russia supersedes US & China as largest malware producer [Электронный ресурс]. – Режим доступа: http://www.pctools.com/news/view/id/197.

2. Researchers invent system to control worms attacking computer networks [Электронный ресурс]. – Режим доступа: http://live.psu.edu/story/22189.

3. Булахов Н.Г. Защита пакетов, передаваемых в сети Ethernet, и её описание как динамической системы / Н.Г. Булахов, В.Я. Хасанов, Б.Н. Пойзнер // Проблемы информационной безопасности государства, общества и личности: материалы 7-й Всероссийской научно-практической конференции, г. Томск, 16–18 февраля 2005 г. – Томск : ИОА СО РАН, 2005. – С. 79–81.

4. Сердюк В. Вы атакованы — защищайтесь! [Электронный ресурс]. – Режим доступа: http://www.bytemag.ru/articles/detail.php?ID=9036.

5. Корт С.С. Методы обнаружения нарушителя [Электронный ресурс]. – Режим доступа: http://www.ssl.stu.neva.ru/sam/

6. Матиас Р. Анализ поведения и эвристические методы выявления вирусов [Электронный ресурс]. – Режим доступа: http://www.osp.ru/lan/2006/10/3474604/

7. Chen W.H. Application of SVM and ANN for intrusion detection / W.H. Chen, Sh.H. Hsu, H.P. Shen [Электронный ресурс]. – Режим доступа: http://dx.doi.org/10.1016/j.cor.2004.03.019.

8. Гудилин О. Проактивность как средство борьбы с вирусами [Электронный ресурс]. – Режим доступа: http://www.viruslist.com/ru/analysis?pubid=189544544.

9. Zou C.C. Code Red Worm Propagation Modeling and Analysis / C.C. Zou, W. Gong, D. Towsley [Электронный ресурс]. – Режим доступа: http://tennis.ecs.umass.edu/~czou/research/codered.pdf

10. Kim J. Measurement and Analysis of Worm Propagation on Internet Network Topology / J. Kim, S. Radhakrishnan, S.K. Dhall [Электронный ресурс]. – Режим доступа: http://ieeexplore.ieee.org/iel5/9617/30391/01401716.pdf

11. Статистическая модель ЦИС, учитывающая возможность пиковых нагрузок / Н.Г. Булахов, Б.Н. Пойзнер, А.Л. Турицин, В.Я. Хасанов // Материалы международной научной конференции «Статистические методы в естественных, гуманитарных и технических науках», Таганрог, апрель 2006 г. – Ч. 3. – Таганрог : Антон, ТРТУ, 2006. – С. 7–11.

Булахов Николай Георгиевич
Аспирант кафедры квантовой электроники и фотоники Радиофизического факультета Томского государственного университета
Тел.: (3822) 41 38 25
Эл. почта: nbooiahov@yandex.ru

N.G. Bulakhov

COMPUTER VIRUSES AND NETWORK WORMS PROPAGATION DETECTION METHODS

Paper introduce statistical model of digital information network applicable for detection of harmful network software propagation

---