© Ю.М. Монахов, 2008

УДК 621.391/519.72

*Ю.М. Монахов*

# *ИСПОЛЬЗОВАНИЕ FARIMA-МОДЕЛИ ДЛЯ ОПИСАНИЯ И ПРЕДСКАЗАНИЯ ПОВЕДЕНИЯ СЕТИ ПЕРЕДАЧИ ДАННЫХ В УСЛОВИЯХ АТАК ТИПА «ОТКАЗ В ОБСЛУЖИВАНИИ»*

*Проведен анализ возможности FARIMA-модели динамики изменения уровня потерь пакетов в сети в качестве основного механизма предсказания и обнаружения DoS-атаки, на основе математического описания сетевой атаки типа «отказ в обслуживании» (DoS-атака) и результатах исследований, показывающих самоподобие сетевого трафика в условиях перегрузок.*

Атака типа «отказ в обслуживании», или *DoS*-атака (*Denial of Service*), представляет собой действие, направленное на приведение жертвы атаки в состояние неспособности выполнять свою работу. В роли жертвы выступают объекты сети передачи данных (СПД). Одним из распространенных методов проведения *DoS*-атаки является перегрузка входящего канала жертвы. Атакующие узлы пытаются загрузить всю полосу пропускания канала пакетами с данными, не несущими никакого смысла, с подставным адресом отправителя. Примером такой атаки служит *ICMP-flood* атака [1].

Положим, что атакуемый узел (сервер) имеет входящий канал с пропускной способностью $C$ бит/с, а пограничный маршрутизатор – входной буфер размером $B$ бит. Ситуация атаки может быть симулирована с помощью модели статистического мультиплексирования трафика от $N$ атакующих узлов, которые могут находиться в двух состояниях: отсылки пакетов (*on*-состояние) и бездействия (*off*-состояние).

Обозначим периоды времени функционирования и бездействия как $t_1$ и $t_2$; $\beta$ и $\alpha$ - соответствующие частоты, то есть $t_1 = \beta^{-1}$ и $t_2 = \alpha^{-1}$. Если источник является активным, то он генерирует $r$ пакетов в секунду. Размер посылаемого пакета в битах обозначим как $L$, а объем полученных пакетов в момент времени $t$ как $Q(t)$. Тогда вероятность перегрузки буфера может быть аппроксимирована формулой [1]

$$P\left(Q\left(t\right) > B\right) = e^{-\gamma B},$$

где

$$\gamma = \left(\frac{\beta}{NrL - C} - \frac{\alpha}{C}\right) \cdot N .$$

Чтобы вызвать перегрузку канала передачи, атакующему необходимо выбрать такие параметры атаки, чтобы значение $\gamma$ было близко к нулю или отрицательное. Поэтому число узлов для осуществления атаки должно удовлетворять неравенству

$$N \geq \frac{C\left(\beta + \alpha\right)}{\alpha r L} .$$

Назовём коэффициентом занятости источника трафика - $\tau$ отношение продолжительности отсылки пакетов источником ко всему периоду функционирования и бездействия

$$\tau = \frac{\beta^{-1}}{\alpha^{-1} + \beta^{-1}}.$$

Тогда неравенство, ограничивающее количество атакующих узлов, можно выразить как $N \geq \frac{C}{r\tau L}$.

В типичных *ICMP-flood* атаках атакующие узлы постоянно находятся в *on*-состоянии, направляя паразитный трафик жертве. В этом случае коэффициент $\tau = 1$ и неравенство для количества атакующих узлов упро-щается до $N \geq \frac{C}{rL}$.

Чтобы атака стала трудно распознаваема, злоумышленник маскирует ее под обычную перегрузку в сети. Для этого ему нужно подобрать достаточно малые значения для параметров $\tau$ и $r$. Так, если злоумышленник выберет значения $\tau = 0{,}05$ и $r = 20$ пак/с, а целью атаки является сервер с каналом пропускной способностью $C = 10$ Мбит/с, то понадобится $1{,}7 \cdot 10^4$ атакующих узлов для проведения атаки.

Кроме того, злоумышленник может выполнять несколько процессов на каждом из узлов атаки, а те, используя фиктивные адреса отправителей, смогут выступать для атакуемого в качестве различных источников трафика. Таким образом, осуществить распределенную *DoS*-атаку, маскируя ее под естественные перегрузки канала, достаточно реально.

Сделанные выводы можно обобщить и на различные типы источников атаки. Пусть у атакующего имеется $M$ различных типов атакующих узлов. Тогда атакующий должен так подобрать параметры атаки, чтобы выполнялось соотношение

$$\sum_{m=1}^{M} N_m r_m L_m \tau_m \geq C.$$

Имея в своём распоряжении достаточное число $N$ подконтрольных узлов, злоумышленник может генерировать трафик, не вызывающий подозрений и аналогичный трафику от обычного пользователя.

Тем не менее, установить факт атаки, а, следовательно, и выбрать соответствующее противодействие злоумышленнику в СПД, все-таки возможно. Будем полагать, что на основании некоторой собранной статистики о динамике состояния сети (например, уровне потерь пакетов) – процессе $\{X_t\}$ – можно дать оценку будущим состояниям изучаемой системы.

Выдвигается гипотеза о том, что трафик СПД в условиях DoS-атаки обладает свойствами персистентности и самоподобия, и, как следствие, имеет медленно убывающую зависимость.

Важнейшим параметром, характеризующим степень самоподобия, является параметр Херста (*Hurst*) - $H$, определяемый для временного ряда $\{X_t\}$ из соотношения

$$\frac{R}{S} = (aN)^H,$$

где $R = max(\{X_t\}) - min(\{X_t\})$ - размах отклонения; $S = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left(x_k - \overline{x}\right)^2}$ - стандартное отклонение, $N$ - число членов временного ряда, $a$ - константа.

Используя значение показателя Херста, выделяют три типа случайных процессов:

1. $0<=H<=0{,}5$ - случайный процесс является антиперсистентным, или эргодическим, рядом, который не обладает самоподобием;
2. $H = 0{,}5$ - полностью случайный ряд, аналогичный случайным смещениям частицы при классическом броуновском движении;

3. $H > 0,5$ - персистентный (самоподдерживающийся) процесс, который обладает длительной памятью и является самоподобным.

Дополнительно следует отметить, что самоподобный процесс часто носит взрывной (*burst*) характер, что выражается в возможности наличия выбросов во время относительно низкой скорости поступления событий. Применительно к трафику самоподобие выражается в неизменности поведения при изменении временных масштабов наблюдения и сохранения склонности к всплескам при усреднении по шкале времени.

Близость параметра Херста к 1 для агрегированного трафика, показанная в [2], указывает на то, что трафик в СПД, находящейся в условиях перегрузок, имеет длительную зависимость от начальных условий (*LRD, long-range dependence*). Такие системы могут быть с достаточной точностью описаны и предсказаны *FARIMA*-моделью (смешанная модель авторегрессии и скользящего среднего) [3].

*FARIMA*($p$, $d$, $q$) - процесс – процесс $\{X_t, t = \ldots -1, 0, 1, \ldots\}$, где параметры $p$ и $q$ - неотрицательные целые числа, $d$ лежит в пределах $-0,5<d< 0,5$.

Процесс описывается разностным уравнением

$$\Phi(B)\Delta^d X_t = \Theta(B) a_t ,$$

где

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p ,$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q ,$$

$B$ – оператор взятия предыдущего значения в множестве $\{X_t\}$, $\Delta^d = (1 - B)^d$ - оператор дробной разности и $\{a_t, t = \ldots -1, 0, 1, \ldots\}$ – белый шум с дисперсией $\sigma^2$.

В нашем случае процесс $\{X_t\}$ представляет собой динамику изменения потерь пакетов в агрегированном трафике изучаемой СПД.

Если полиномы $\Phi(z)$ и $\Theta(z)$ не имеют общих корней, и все корни лежат вне единичной окружности, то процесс $\{X_t\}$ является стационарным и обратимым. Если параметр $d$ удовлетворяет неравенству $0<d<0,5$ , то описываемая система имеет длительную зависимость с параметром Херста $H = d+0,5$.

Применение *FARIMA* - модели для процесса является более трудоёмким, чем его моделирование с помощью *ARMA* - процесса [3]. В то же время обсчёт *FARIMA* - процесса можно разбить на две составляющие – генерацию дробно-разностного шума *FARIMA(0,d,0)* и моделирование *ARMA(p,q)* - составляющей.

Для *FARIMA* - процесса мы можем записать

$$W_t = \Delta^d X_t ,$$

где

$$W_t = \frac{\Theta(B)}{\Phi(B)} a_t .$$

Тогда, если можно определить величину степени различия $d$ и значения для оператора дробной размерности, то $\{W_t\}$ - дробно-разностная составляющая $\{X_t\}$ - может быть представлена *ARMA(p,q)* - моделью.

Как показано в [3], показатель степени различия и параметр Херста для *LRD*-процессов связаны соотношением $H = d + 0,5$. Таким образом, вычислив параметр Херста для рассматриваемой системы, применяя *R/S*-статистику или какой-либо другой метод, можно найти и показатель степени различия $d$.

Вычисление $\{W_t\}$ из $\{X_t\}$ требует конечного числа шагов аппроксимации

$$\Delta^d\left(X_t-u\right)=\sum_{k=0}^{\infty}\left(-B\right)^k\left(X_t-u\right)=$$

$$=\sum_{j=0}^{\infty}\overline{\omega_j}\left(X_{t-j}-u\right),$$

где

$$\overline{\omega_j}=\frac{\left(-1\right)^j\Gamma\left(1+d\right)}{\Gamma\left(1+j\right)\Gamma\left(1+d-j\right)}.$$

Как видно, формула вычисления значения оператора дробной размерности включает в себя значения $X_0$, $X_{-1}$, $X_{-2}$, которые не могут быть замерены системой обнаружения в сети. Вследствие этого их необходимо заменить средним значением $u$.

Теперь становится возможным вычислить значения параметров $p$ и $q$ для *ARMA* - модели, проанализировав составляющую $\{W_t\}$. Это можно сделать, применив информационный критерий Акаике (*Akaike Information Criterion*) [4] или информационный критерий Байеса (*Bayesian Information Criterion*). Появляется возможность вычислить все параметры $\phi_1,\phi_2,\ldots,\phi_p$ и $\theta_1,\theta_2,\ldots\theta_q$, а также $\sigma^2$.

Так как $\{a_t,\ t = \ldots-1, 0, 1, \ldots\}$ – белый шум с постоянной дисперсией $\sigma^2$, то теперь, вычислив все параметры *FARIMA* - модели, мы можем предсказать значение $X_t$:

$$X_t=\Phi^{-1}\left(B\right)\cdot\Theta\left(B\right)Y_t,$$

где $Y_t=\Delta^d a_t$

Таким образом, моделирование процессов, происходящих в СПД, с помощью *FARIMA*-модели, учитывая их самоподобную природу и долговременную зависимость от начальных условий, предоставляет возможность оценить будущее состояние системы на основе её текущих параметров.

Система обнаружения атак на основе анализа текущей динамики изменения уровня потери пакетов в сети, построенная с применением *FARIMA*-модели (в качестве изучаемого процесса которой выступает изменение уровня потерь), может оценить угрозу реализации распределённой *DoS*-атаки в сети. В данном случае система обнаружения атак по росту предсказанного значения $X_t$ может сделать вывод о проведении в СПД распределённой *DoS*-атаки.

***СПИСОК ЛИТЕРАТУРЫ***


1. *Huang Q.* «Analysis of a New Form of Distributed Denial of Service Atttack». Dep. Of Electrical Engineering, Princeton University.

2. *Fekete A. and Vattay G.* «Self-similarity in bottleneck buffers». Proceedings of Globecom 2001, December 2001.

3. *Fei Xue.* «Modeling and Predicting Long-range Dependent Traffic with FARIMA Processes». Dept. of Information EngineeringThe Chinese University of Hong Kong. http://www.ensc.sfu.ca/~ljilja/cnl/papers/mflrd.ps.

4. *Bozdogan H.* «Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions». ТИАБ


## *Коротко об авторе*


*Монахов Ю.М.* – аспирант кафедры информационных систем и информационного менеджмента Владимирского государственного университета.



Рецензент канд. техн. наук, доцент *Д.В. Александров,* Владимирский государственный университет.