УДК 004.942

МОДЕЛИРОВАНИЕ СЕГМЕНТОВ АДРЕСНОГО ПРОСТРАНСТВА СЕТИ ИНТЕРНЕТ

© 2018 К.Е. Климентьев

Самарский национальный исследовательский университет имени академика С.П. Королёва

Статья поступила в редакцию 12.12.2018

Рассматриваются принципы построения моделей адресного пространства сети Интернет в условиях недоступности отдельных узлов и групп узлов. В качестве математической модели предлагается модель случайных разбиений.

Ключевые слова: самовоспроизводящийся объект, сетевой червь, эпидемия, сеть, узел, адресное пространство, случайное разбиение, диаграмма Юнга, предельная кривая, аппроксимация.

ВВЕДЕНИЕ

В статье рассматриваются вопросы моделирования адресных пространств, неоднородных с точки зрения доступности локаций для заражения. Актуальность обусловлена необходимостью исследования развития в таких пространствах эпидемий SI-типа.

Понятие эпидемий SI-типа рассмотрено, например, в [1]. Такое обозначение получили модели эпидемий, в развитии которых участвуют сущности двух типов: S – пригодные для инфицирования; I – инфицированные. Развитие заключается в переносе свойства инфицированности с сущностей типа I на сущности типа S со средней скоростью β попыток инфицирования в единицу времени, причем обратный перенос невозможен. Очевидно, важным параметром поведения SI-моделей является алгоритм поиска сущностями типа I сущностей типа S в некотором адресном пространстве, в котором каждой сущности сопоставлен некоторый (например, числовой) идентификатор. Рассмотрим наиболее важные частные случаи.

Случай 1: беспрепятственное размножение. Считается, что перечень идентификаторов сущностей, доступных для инфицирования (то есть, сущностей типа S), в любой момент достоверно известен. В классической работе [2] такой перечень назван «хитлистом» (англ. hitlist – список целей). Очевидно, в этом случае развитие эпидемии происходит по экспоненциальному закону $I=I_0(1+\beta)^t$ и завершается в максимально короткие сроки. Например, если эпидемия начинается с единственного заражения $I_0=1$ и сетевой червь способен выполнять одно инфицирование в секунду (то есть $\beta=1$), то абсолютно все узлы сети Интернет, имеющие 32-битовые адреса в рамках IPv4, будут инфицированы всего за полминуты. Алгоритм поведения такого гипотетического «блицкриг»-червя рассмотрен, например, в [3], – кратко охарактеризуем его.

Климентьев Константин Евгеньевич, кандидат технических наук, доцент кафедры информационных систем и технологий. E-mail: climentieff@ro.ru

Пусть адресное пространство представляет собой непрерывный массив одинаковых ячеек, каждая из которых содержит либо сущность типа I, либо типа S. Все ячейки последовательно пронумерованы, начиная с 1. В момент времени t=0 все ячейки, начиная с адреса 1 и до адреса I_0, считаются занятыми сущностями типа I. Такого положения дел всегда можно достичь перенумерацией ячеек. Время считается дискретным, акт «заражения» происходит мгновенно. Сущности на каждом шаге «срабатывают» в порядке увеличения занятых ими адресов, при этом занятие свободных ячеек так же производится в направлении от младших адресов к старшим. Каждая сущность типа I непосредственно перед актом размножения вычисляет область адресного пространства, в котором будут размещены ее «потомки», причем эта область не пересекается с аналогичными областями других экземпляров инфицирующих сущностей. Для сущности типа I, расположенной по адресу n, начальный адрес области адресного пространства, имеющей длину b и предназначенной для последующего занятия «потомками» этой сущности в произвольный момент времени t, может быть вычислен как $n_t = I_0(1+\beta)^{t-1} + (n-1)\beta + 1$.

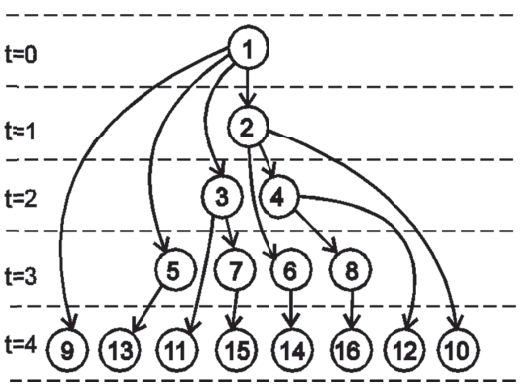


Рис. 1. Дерево потомков при послойном заполнении адресного пространства в условиях $I_0=1$, b=1

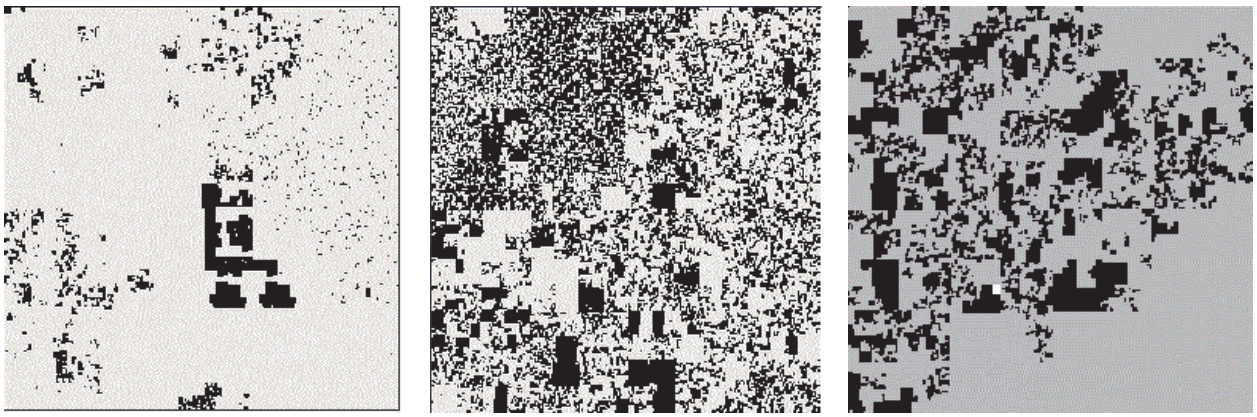
Пример заполнения адресного пространства с использованием этого алгоритма приведен на рис. 1.

Случай 2: размножение в частично недоступном пространстве. Очевидно, если «хит-лист» неизвестен, вышерассмотренный алгоритм работы потерпит неудачу. Это связано с тем обстоятельством, что значительная часть узлов сети Интернет в любой момент времени либо временно отключена, либо задействована во внутренних сетях без доступа с сети Интернет, либо вообще не используется. В 2012 г. (уже после официального «исчерпания» свободных адресов!) исследовательский ботнет Carna, просканировав все адресное пространство Интернет, обнаружил лишь 1.3 млрд. доступных адресов IPv4, что составило 30% от 4.3 млрд [4]. Более современные результаты сканирования доступны по адресу https://scans.io. Недоступность части адресов существенно влияет на любые попытки «вирусного» сканирования Интернета, выполняемого как во вредоносных целях (см., например, описание алгоритма работы сетевого червя Lovesan [5]), так и с «благими намерениями».

Автором настоящей статьи при помощи штатной утилиты ping, доступной как в Windows, так и в клонах UNIX, было проведено сканирование доступности IP-адресов в 51 наугад выбранных сегментах сети Интернет размером 65536 адресов каждый, в равных долях принадлежащих различным классам сетей, а именно [6]: класса A в диапазоне 0.0.0.0¸127.255.255.255, класса B в диапазоне 128.0.0.0¸191.255.255.255 и класса C в диапазоне 192.0.0.0¸225.255.255.255. Соответственно, сканируемые сегменты представляли собой либо целиком какие-либо сети, либо части сетей, либо объединения сетей, принадлежащих разным «хозяевам». Примеры просканированных сегментов см. на рис. 2,а и 2,б. В общем случае структура этих сегментов постоянно меняется, но на временных интервалах, в течение которых производится сканирование (например, со стороны сетевого червя), их можно считать статичными.

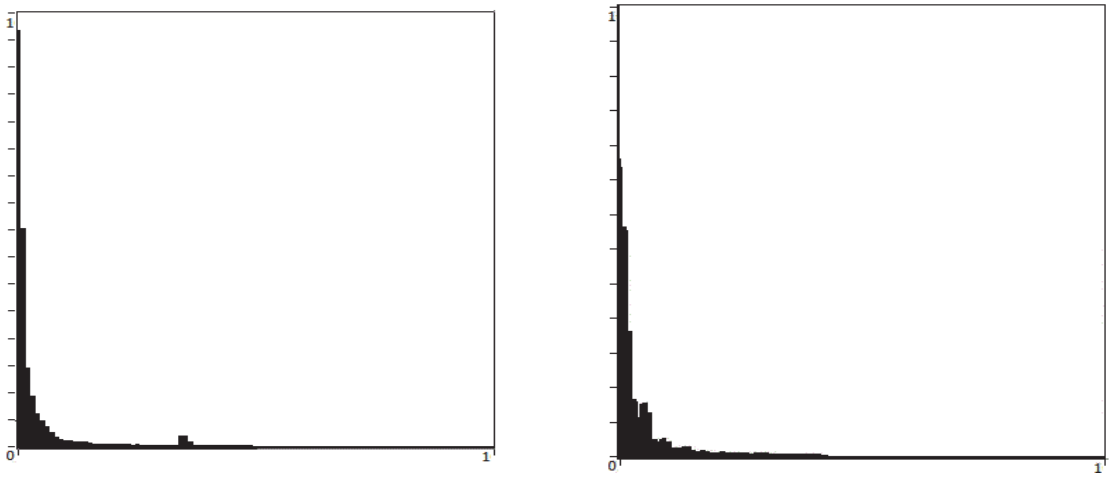
Задача исследования. Была поставлена задача построения моделей, адекватно (в смысле статистического распределения доступных и недоступных узлов) отражающих структуру сегментов сетей.

Гистограммы эмпирических распределений интервалов между адресами двух соседних «доступных» узлов, а так же, соответственно, между адресами «недоступных» узлов, полученные в результате сканирования, приведены на рис. 3.



а) Слабо заполненный сегмент б) Умеренно заполненный сегмент в) Модель, |M|/|K|=3/7

Рис. 2. Карты доступности сегментов сетей (черным цветов – «живые» адреса)



а) Интервалы между доступными адресами б) Интервалы между недоступными адресами

Рис. 3. Гистограммы распределений длин интервалов

В принципе, их достаточно для моделирования случайных сегментов адресного пространства сети Интернет, «похожих на правду» (см., например, рис. 2,в). Тем не менее, поскольку выборка из 51 сегмента не может быть признана репрезентативной, встает задача обоснования адекватности моделей, сгенерированных на основе эмпирически полученных распределений.

Математическая модель адресного пространства. Пусть размер сегмента адресного пространства составляет |N|=|M|+|K|, де |M| количество «недоступных», а |K| - количество «доступных» ячеек в интервале, содержащем ячейки одного типа. В свою очередь, |M| и |K| могут быть представлены в виде своих разбиений $|M|=m_1+m_2+m_3+\ldots$ и $|K|=k_1+k_2+k_3+\ldots$ на слагаемые, где каждому из слагаемых сопоставляется непрерывная область адресного пространства, содержащая только «недоступные» или только «доступные» ячейки. Например, |N|=12, |M|=2+1+1+2 и |K|=1+3+1+1 (см. рис. 4).

Таким образом, любая конкретная конфигурация сегмента адресного пространства может быть представлена в виде совокупности двух разбиений, тем или иным образом (например, случайно) выбранных из множества всевозможных разбиений чисел |M| и |K|. Итак, математическая модель сегмента адресного пространства сети Интернет может быть определена следующим образом:

- $\{(m_i,k_i)\}$ – множество пар элементов множеств $\{m_i\}$ и $\{k_i\}$;
- $\{m_i\}$ – множество целых чисел, являющихся реализациями случайной величины, распределенной по закону F_M, приближенно соответствующему рис. 3,а;
- $\{k_i\}$ – множество целых чисел, являющихся реализациями случайной величины, распределенной по закону F_K, приближенно соответствующему рис. 3,б;
- |M|, |K|, |N| - мощности множеств $\{m_i\}$, $\{k_i\}$ и $\{(m_i,k_i)\}$ соответственно, причем |N|=|M|+|K|.

Осталось более точно и адекватно определить вид распределений F_M и F_K.

Распределение элементов случайных разбиений. Очень удобной формой представления разбиений являются диаграммы Юнга [7]. Например, для рис. 4 без разделения на «доступные» и «недоступны» элементы диаграмма Юнга (точнее, ее «французская» разновидность) будет иметь вид, изображенный на рис. 5,а.

Показано (см., например [8]), что при N®¥ множество форм всевозможных случайных диаграмм Юнга сходится к некоторой «типичной» кривой W(x), обладающей следующими свойствами: 1) монотонным убыванием; 2) симметричностью относительно основной диагонали I-го квадранта. Длина первой строки и первого столбца «типичной» диаграммы Юнга сходится к $2\sqrt{N}$. Вид этой кривой, изображенной на рис. 5,б, можно визуально сравнить с рис. 3. Таким образом, при соблюдении условия нормировки $\int_{-\infty}^{+\infty}\Omega(x)dx=1$ кривая W(x) может рассматриваться как плотность распределения вероятности длин интервалов «доступных» и «недоступных» адресов.

Аппроксимация кривой Ω(x) была предложена в работе [9]: $e^{-\pi x/\sqrt{6}}+e^{-\pi y/\sqrt{6}}=1$, от-

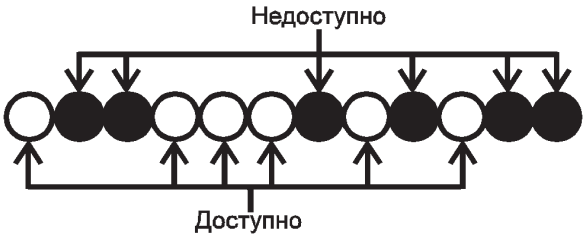
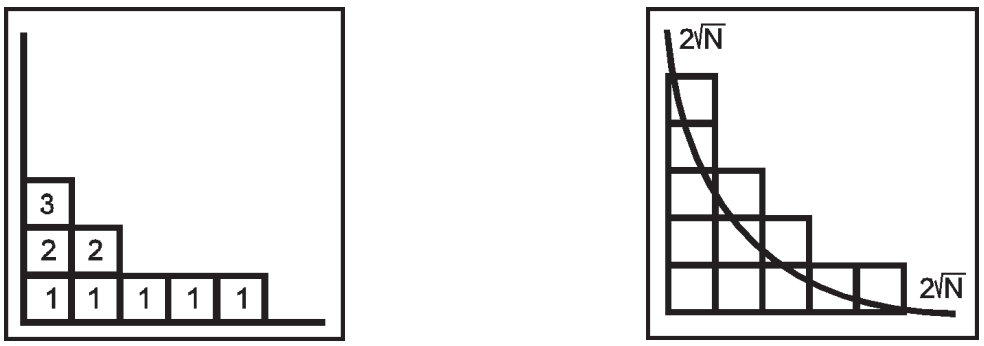


Рис. 4. Совокупность числовых разбиений как модель сегмента



а) Пример диаграммы б) «Типичная» форма случайной диаграммы

Рис. 5. Диаграммы Юнга для числовых разбиений

куда $\Omega(x) = -\frac{\sqrt{6}}{\pi}\left(1 - e^{-\pi x/\sqrt{6}}\right)$.

Точная формула кривой Ω(x), повернутой на π/4 против часовой стрелки вокруг начала координат, была получена в работе [10].

В параметрической форме:

$$\left\{\begin{array}{l} x = y + 2\cos\theta \\ y = \frac{2}{\pi}(\sin\theta - \theta\cos\theta) \\ 0 \le \theta \le \pi \end{array}\right\}.$$

В обычной форме:

$$\Omega(x) = \left\{\begin{array}{l} \frac{2}{\pi}\left(x\arcsin\left(\frac{x}{2}\right) + \sqrt{4 - x^2}\right), |x| \le 2 \\ |x|, else. \end{array}\right\}.$$

РЕЗУЛЬТАТЫ И ВЫВОДЫ

Для вышеприведенных аппроксимаций условие нормировки выполняется, и они могут быть использованы для моделирования случайных величин, характеризующих разбиение множества узлов сегмента сети Интернет на «доступные» и «недоступные» узлы. Пример результата моделирования для отношения |M|/|K|=3/7 приведен на рис. 2,в. Результаты настоящей работы предназначены для использования в системе моделирования размножения самовоспроизводящихся сущностей (см. [11]) с целью исследования алгоритмов эффективного сканирования адресных пространств.

СПИСОК ЛИТЕРАТУРЫ

1. *Hethcote W.* The Mathematics of Infectious Diseases // SIAM REVIEW, 2000. - Vol. 42, No. 4, pp. 599–653.
2. *Weaver N.C.* Warhol Worms: The potential for very fast internet plagues? - 2001. - Режим доступа: http://www1.icsi.berkeley.edu/nweaver/papers/warhol/warhol.html.
3. *Климентьев К.Е.* Оптимальное сканирование адресного пространства во время эпидеий SI-типа // Материалы международной научно-технической конференции ПИТ-2018. – Самара: Изд-во СНЦ РАН, 2018. - с. 290-292.
4. *Trapickin R.* Who is scanning the Internet? // Seminars FI/IITM SS 15, Munchen 2015. – pp. 81-88.
5. *Климентьев К.Е.* Компьютерные вирусы и антивирусы: взгляд программиста. – М.: ДМК-Пресс, 2013. – 656 с.
6. *Мамаев М., Петренко С.* Технологии защиты информации в Интернете. Специальный справочник. – СПб.: Питер, 2002. – 848 с.
7. *Смирнов Е.Ю.* Диаграммы Юнга, плоские разбиения и знакочередующиеся матрицы. – М: МЦНМО, 2014 – 62 с.
8. *Буфетов А.И., Житлухин М.В., Козин Н.Е.* Диаграммы Юнга и их предельная форма. – М: МЦНМО, 2013. – 56 с.
9. *Temperley H.* Statistical mechanics and the partition of numbers. The form of the crystal surfaces // Proc. Cambridge Philos. Soc. 48, 1952. - pp. 683-697.
10. *Вершик А.М., Керов С.В.* Асимптотика меры Планшереля симметрической группы и предельная форма таблиц Юнга. // ДАН СССР, 233, вып.6, 1977. - С. 1024–1027.
11. *Климентьев К.Е.* Мультиагентное моделирование процессов распространения и взаимодействия “инфицирующих” сущностей // Программные продукты и системы. - Тверь, 2018. - 1(31) - с. 744-748.

MODELING SEGMENTS OF THE INTERNET ADDRESS SPACE

Samara National Research University named after Academician S.P. Korolyov

Random partitions as a model of partially unavailable Internet address space is discussed.

Keywords: selfreproductive object, networm, spreading, epidemics, disease, network, node, address space, random partition, Young tableau, limit shape, approximation.

Konstantin Klimentiev, Candidate of Technics, Associate Professor at the Information Systems and Technologies Department. E-mail: climentieff@ro.ru