УДК 519.245:004.056

МОДЕЛИРОВАНИЕ ЗАДЕРЖАННОГО РАСПРОСТРАНЕНИЯ ВРЕДОНОСНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ МЕТОДОМ МОНТЕ-КАРЛО

Назаров Зухриддин Акбар угли,
доцент кафедры "”Математика и естественные науки" государственного университета Cyber University, PhD
zuhrov13@gmail.com
Исмоилова Муниса Ориф кизи,
преподаватель кафедры "”Цифровые технологии"” университета Perfect University
munisaismoilova9718@gmail.com

Аннотация. *В данной статье представлен новый и эффективный подход к имитационному моделированию методом Монте-Карло для ветвящихся процессов с задержкой на периодах размножения. Хотя теория ветвящихся процессов широко применяется в биологии, генетике и эпидемиологии, в данной работе она адаптируется к области кибербезопасности, в частности, для моделирования процесса задержанного распространения вредоносных программ в сети. В модели каждая «частица» представляет собой заражённый узел, её потомки - новые узлы, заражённые от неё, а период задержки отражает скрытое состояние вредоносной программы до момента её активации.*

В статье сначала излагаются математические основы ветвящихся процессов с задержкой на периодах размножения, вероятностные распределения и рекуррентное выражение для числа частиц $n-го$ поколения. Затем предлагается алгоритм Монте-Карло для моделирования данного процесса.

Ключевые слова: *моделирование Монте-Карло, ветвящийся процесс, задержка размножения, вредоносная программа, кибербезопасность, распространение в сети.*

MONTE CARLO SIMULATION OF DELAYED MALWARE PROPAGATION

Abstract. *This paper presents a new and efficient Monte Carlo simulation approach for branching processes with delay at reproduction periods. Although the theory of branching processes has been widely applied in biology, genetics, and epidemiology, this work extends it to the field of cybersecurity, specifically for modeling the delayed propagation of malware across a network. In the model, each "particle" represents an infected node, its offspring represent new nodes infected from it, and the delay period reflects the dormant state of the malware prior to activation.*

The paper first presents the mathematical foundations of branching processes with delay at reproduction periods, their probability distributions, and the recursive expression for the number of particles in the $n-th$ generation. A Monte Carlo algorithm for simulating this process is then proposed.

Keywords: *Monte Carlo simulation, branching process, delay at reproduction, malware, cybersecurity, network propagation.*

ZARARLI DASTURNING KECHIKTIRILGAN TARQALISHINI MONTE-KARLO USULIDA MODELLASHTIRISH

Annotatsiya. *Ushbu maqolada ko‘payish davrlarida kechikkan holdagi tarmoqlanuvchi jarayonlar uchun yangi va samarali Monte-Karlo simulyatsiya yondashuvi taqdim etiladi. Tarmoqlanuvchi jarayonlar nazariyasi biologiya, genetika va epidemiologiyada keng qo‘llanilgan bo‘lsa-da, ushbu ishda u kiberxavfsizlik sohasiga, xususan, zararli dasturlarning tarmoqda kechikib tarqalish jarayonini modellashtirish uchun tatbiq etiladi. Modelda har bir "zarracha" zararlangan tugunni, uning avlodlari esa undan yuqtirilgan yangi tugunlarni ifodalaydi, kechikish davri esa zararli dasturning faollashishidan oldingi yashirin holatini aks ettiradi.*

Maqolada dastlab ko‘payish davrlarida kechikishga ega tarmoqlanuvchi jarayonlarning matematik asoslari, ehtimollik taqsimotlari va $n-avlod$ zarrachalari sonining rekursiv ifodasi keltiriladi. So‘ngra ushbu jarayonni simulyatsiya qilish uchun Monte-Karlo algoritmi taklif etiladi.

Kalit so'zlar: *Monte-Karlo simulyatsiyasi, tarmoqlanuvchi jarayon, ko'payishda kechikish, zararli dastur, kiberxavfsizlik, tarmoqda tarqalish.*

Введение. Теория ветвящихся процессов представляет собой хорошо разработанную область прикладной теории вероятностей, описывающую системы, в которых сущности-частицы, особи, клетки или, в более широком смысле, активные агенты внутри сети-существуют в течение определённого периода времени, а затем заменяются случайным числом подобных сущностей. Начиная с классического исследования Ф. Гальтона о вымирании фамилий среди британской аристократии, ветвящиеся процессы обеспечивают строгий математический аппарат для описания явлений роста, разворачивающихся через последовательные, независимые события воспроизводства. Со временем эта теория превратилась в зрелую и активно развивающуюся ветвь прикладной теории вероятностей, внёсшую существенный вклад в молекулярную биологию, клеточную биологию, биологию развития, иммунологию, эволюционную теорию, экологию и медицину. В этих областях модели ветвящихся процессов помогли исследователям интерпретировать противоречащие интуиции экспериментальные наблюдения, разрабатывать новые эксперименты и клинические протоколы, а также формулировать прогнозы, которые впоследствии были подтверждены в реальных условиях.

В классическом ветвящемся процессе Беллмана–Харриса единственная частица-прародитель рождается в момент времени $t = 0$ и живёт в течение случайного времени τ, распределение которого задаётся функцией распределения $G(\tau)$. В момент гибели частица порождает случайное число потомков в соответствии с производящей функцией вероятностей $f(s)$, причём каждый из этих потомков ведёт себя независимо от остальных и от прародителя, следуя тому же вероятностному закону.

Если $Z(t)$ обозначает число частиц, живых в момент времени t, то результирующий стохастический процесс

$$\{Z(t),\ t \geq 0\}$$

в общем случае не является марковским; тем не менее, ряд важных частных случаев, в первую очередь, процесс Гальтона–Ватсона и возраст-зависимый процесс с экспоненциально распределёнными временами жизни восстанавливает марковское поведение при упрощающих предположениях относительно распределения времени жизни.

Естественное расширение этой классической схемы возникает, когда время жизни каждой частицы ограничивается дискретными, целочисленными единицами времени, а само воспроизводство подвергается дополнительной задержке: вместо немедленного размножения частица может оставаться неактивной в течение одного или нескольких периодов после своего появления, прежде чем произвести потомство. Такая конструкция приводит к ветвящемуся процессу с задержкой периодов размножения (ВПЗПР), в котором совместное распределение вероятностей числа потомков X и периода задержки M совместно определяет эволюцию численности популяции $Z(n)$ на протяжении последовательных поколений. Поскольку численность популяции в поколении n зависит от вкладов всех предшествующих поколений, а не только от непосредственно предшествующего, возникающая динамика оказывается существенно более богатой, чем в стандартном процессе Гальтона–Ватсона, а замкнутые аналитические выражения для таких величин, как распределение $Z(n)$, в общем случае получить затруднительно. Ввиду этого моделирование методом Монте-Карло представляет собой естественный и вычислительно реализуемый инструмент для изучения подобных процессов, позволяющий генерировать выборочные траектории $Z(n)$ непосредственно из лежащей в основе вероятностной структуры без необходимости получения явного аналитического решения.

Хотя первоначальная мотивация для изучения моделей типа ВПЗПР носила преимущественно биологический характер, например, анализ распространения генов и задержанного размножения в клеточных популяциях, та же лежащая в её основе математическая структура способна описывать значительно более широкий класс явлений задержанного распространения. В частности, распространение вредоносного программного обеспечения по компьютерной сети демонстрирует тесно связанное с этим ветвящееся поведение: заражённый узел, рассматриваемый здесь как «частица», может оставаться в состоянии покоя в течение случайного числа дискретных периодов, прежде чем активироваться и, после активации, заразить случайное число новых узлов, которые, в свою очередь, составляют его «потомство». Такая модель поведения характерна для скрытного

вредоносного программного обеспечения (ПО), самораспространяющихся червей и ботнетов, использующих задержанные, «спящие» или триггерные стратегии активации, а не немедленное распространение сразу после заражения. Значительная часть существующей литературы по распространению червей и вредоносного ПО опирается на эпидемиологические компартментные модели, такие как формулировки «восприимчивый–инфицированный–выздоровевший» (SIR) или «восприимчивый–инфицированный–восприимчивый» (SIS), либо на стохастические и аналитические подходы в непрерывном времени. Хотя такие модели эффективны для описания агрегированной динамики заражения, они не отражают явным образом дискретную, структурированную по поколениям задержку между моментом заражения и моментом, когда заражённый узел начинает активно распространять вредоносное ПО далее. Переосмысление распространения вредоносного ПО через призму ВПЗПР предлагает, таким образом, взаимодополняющую и в ряде отношений более структурно достоверную перспективу: период задержки M естественным образом отождествляется с латентной, или спящей, фазой вредоносного ПО до активации, тогда как число потомков X представляет собой количество новых заражённых узлов, возникающих при каждом событии активации, причём эти две величины могут быть как независимыми, так и статистически зависимыми, в зависимости от требований лежащего в основе механизма заражения.

Опираясь на данную переинтерпретацию, настоящая статья разрабатывает подход к моделированию методом Монте-Карло в рамках ВПЗПР, специально адаптированный к задаче кибербезопасности, связанной с задержанным распространением вредоносного программного обеспечения. Помимо моделирования отдельных траекторий распространения численности заражённых узлов $Z(n)$, предлагаемый подход уделяет особое внимание оценке надёжности получаемых оценок методом Монте-Карло: анализируется сходимость смоделированных оценок $Z(n)$ к их теоретическим математическим ожиданиям, а точность оценивания дополнительно характеризуется посредством дисперсии и соответствующего доверительного интервала. Такая количественная оценка надёжности оценивания, отсутствующая в исходной формулировке модели ВПЗПР, призвана повысить практическую применимость предлагаемого подхода моделирования для специалистов по кибербезопасности, стремящихся спрогнозировать траекторию распространения вредоносного ПО и спланировать соответствующие превентивные меры.

Анализ литературных источников. Теоретические основы модели ВПЗПР были заложены в более ранних работах. Исходная формулировка ВПЗПР, включая её вероятностную структуру и базовые свойства, такие как рекуррентное соотношение для математического ожидания численности популяции по поколениям, была предложена в [2], тогда как алгоритм моделирования методом Монте-Карло для генерации выборочных траекторий $Z(n)$ в рамках данной модели был предложен в [1]. Более широкий теоретический контекст ветвящихся процессов, включая их биологические интерпретации, а также классические формулировки Беллмана–Харриса и Гальтона–Ватсона, всесторонне рассматривается в [3], а представление общих ветвящихся процессов в виде дискретных случайных деревьев разработано в [4]. Использованная в настоящей статье методология Монте-Карло опирается на общие принципы моделирования методом Монте-Карло для прикладных научных задач, описанные в [5], а также на более широкие методы моделирования и оценивания, включая снижение дисперсии и построение доверительных интервалов, представленные в [6].

Что касается кибербезопасности, значительный объём исследований был посвящён моделированию распространения вредоносного ПО и червей, однако преимущественно в рамках подходов, отличных от ветвящихся процессов. Кефарт и Уайт [9] одними из первых адаптировали эпидемиологические компартментные модели к распространению компьютерных вирусов, представив распространение заражения на ориентированном графе контактов между хостами. Развивая эту эпидемиологическую традицию, Цзоу, Гонг и Тоусли [7] предложили модель распространения червя Code Red, учитывающую влияние человеческих контрмер и установки исправлений на кривую заражения, а впоследствии расширили эту модель в работе [8], обеспечив возможность мониторинга и раннего обнаружения интернет-червей посредством наблюдения за сканирующим трафиком, поступающим в неиспользуемое адресное пространство. Более поздние работы продолжили совершенствовать эпидемиологические и стохастические подходы к распространению вредоносного ПО в специализированных сетевых условиях: Чжу и Хуан [10] разработали модель распространения для кластерных беспроводных сенсорных сетей, учитывающую радиус связи узлов, плотность узлов и игру «атака–защита» между вредоносным ПО и защитными системами, тогда как Беномар и др. [11] предложили агентную стохастико-геометрическую модель

распространения вредоносного ПО в сетях типа «устройство-устройство», явным образом моделирующую случайное время передачи событий между соседними устройствами.

В совокупности рассмотренная литература демонстрирует как зрелость компартментных и стохастико-геометрических подходов к моделированию распространения вредоносного ПО, так и, вместе с тем, относительное отсутствие дискретных, структурированных по поколениям механизмов задержки того рода, который отражён в рамках модели ВПЗПР. В частности, ни одна из рассмотренных выше моделей распространения вредоносного ПО не представляет явным образом период покоя заражённого узла в виде дискретной случайной задержки, определяющей начало дальнейшего распространения, и ни одна из них не предлагает оценку методом Монте-Карло, сходимость и точность которой анализировались бы относительно известного теоретического математического ожидания, как это сделано для модели ветвящегося процесса в работах [1] и [2]. Этот пробел мотивирует переинтерпретацию модели ВПЗПР, разработанную в настоящей статье, наряду с сопутствующим анализом сходимости и дисперсии соответствующей оценки методом Монте-Карло.

Методы решения. Данный класс стохастических процессов мы называем ВПЗПР, переинтерпретируя его здесь в контексте динамики заражения сети: каждая частица представляет собой заражённый узел, а задержка при воспроизводстве соответствует интервалу покоя, в течение которого заражённый узел остаётся неактивным, прежде чем распространить заражение далее. Рисунок 1 иллюстрирует одну из реализаций данного процесса, где один из заражённых узлов, порождённых во втором поколении, остаётся неактивным в течение двух дополнительных периодов, прежде чем активироваться и заразить новые узлы.

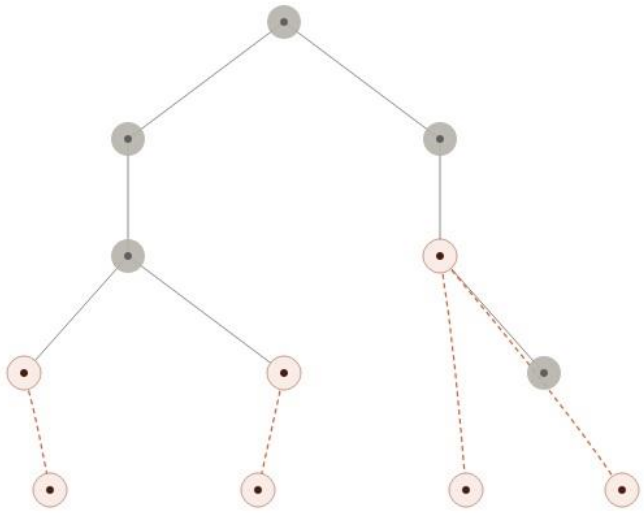


Рисунок 1. Пример ВПЗПР

(---------- немедленное заражение/продолжение; ---------- заражение после задержки.)

Моделирование ВПЗПР. Предположим, что дискретная случайная величина X обозначает число новых заражённых узлов, порождаемых заражённым узлом, а его период покоя представляет собой неотрицательную дискретную случайную величину, обозначаемую M. Тогда совместная функция вероятности для ВПЗПР задаётся функцией

$$f_{X,M}(x,m) := \mathsf{P}(X = x, M = m), \;\; x \in \mathsf{N}_0, \;\; m \in \mathsf{N}, \tag{1}$$

где $\mathsf{N}_0 := \{0\} \cup \mathsf{N}$ и $\mathsf{N} = \{1,2,3,...\}$.

Пример. Предположим, что репродуктивное поведение заражённого узла в ВПЗПР задаётся следующей совместной функцией вероятности.

$M \setminus X$	**0**	**1**	**2**
1	0.08	0.12	0.15
2	0.1	0.1	0.05
3	0.05	0.15	0.2

Мы получаем следующие результаты:

$$\mathsf{E}[X] = 0 \cdot (0.08 + 0.1 + 0.05) + 1 \cdot (0.12 + 0.1 + 0.15) + 2 \cdot (0.15 + 0.05 + 0.2) = 1.17,$$

$$\mathsf{E}[M] = 1\cdot(0.08+0.12+0.15)+2\cdot(0.1+0.1+0.05)+3\cdot(0.05+0.15+0.2) = 2.05,$$

$$\mathsf{E}[X^2] = 0^2\cdot(0.08+0.1+0.05)+1^2\cdot(0.12+0.1+0.15)+2^2\cdot(0.15+0.05+0.2) = 1.97,$$

$$\mathsf{E}[M^2] = 1^2\cdot(0.08+0.12+0.15)+2^2\cdot(0.1+0.1+0.05)+3^2\cdot(0.05+0.15+0.2) = 4.95,$$

$$\mathsf{Var}[X] = \mathsf{E}[X^2]-\Big[\mathsf{E}[X]\Big]^2 = 1.97-1.17^2 = 0.6011,$$

$$\mathsf{Var}[M] = \mathsf{E}[M^2]-\Big[\mathsf{E}[M]\Big]^2 = 4.95-2.05^2 = 0.7475,$$

$$\mathsf{E}[XM] = 1\cdot1\cdot0.12+2\cdot1\cdot0.15+1\cdot2\cdot0.1+2\cdot2\cdot0.05+1\cdot3\cdot0.15+2\cdot3\cdot0.2 = 2.47.$$

Теперь вычислим ковариацию и коэффициент корреляции:

$$\mathsf{Cov}(X,M) = \mathsf{E}[XM]-\mathsf{E}[X]\cdot\mathsf{E}[M] = 2.47-1.17\cdot2.05 = 0.0715,$$

$$\rho(X,M) = \frac{\mathsf{Cov}(X,M)}{\sqrt{\mathsf{Var}[X]\cdot\mathsf{Var}[M]}} = \frac{0.0715}{\sqrt{0.6011\cdot0.7475}} \approx 0.11.$$

В данном примере случайные величины X и M являются зависимыми. Коэффициент корреляции между X и M составляет приблизительно 0.11.

Поскольку формирование численности заражённых узлов в периоде n требует учёта вкладов от всех предшествующих периодов $(n-1),(n-2),\ldots,1,0,$ предположим, что неотрицательные целочисленные случайные величины $X_{ij}^{(n)},\ i,j\in\mathsf{N}$ подчиняются функции вероятности (1) и обозначают число новых заражённых узлов, порождаемых $j-$заражённым узлом, где сам этот узел был активирован в периоде $n-i$ и производит своё потомство в периоде n. Иными словами, данный заражённый узел оставался неактивным в течение i периодов, прежде чем размножиться.

Если обозначить через $Z(n)$ число заражённых узлов в поколении n, то мы получаем:

$$Z(n) = \sum_{i=1}^{n}\sum_{j=1}^{Z(n-i)} X_{ij}^{(n)}. \tag{2}$$

Например, на приведённой выше диаграмме мы имеем:

$$Z(4) = (X_{11}^{(4)}+X_{12}^{(4)}+X_{13}^{(4)})+(X_{21}^{(4)}+X_{22}^{(4)})+(X_{31}^{(4)}+X_{32}^{(4)})+X_{41}^{(4)}.$$

Здесь первая группа $(i=1)$ соответствует трём заражённым узлам, рождённым в периоде 3, из которых два размножаются сразу в периоде 4, а третий остаётся неактивным и не даёт вклада; вторая группа $(i=2)$ соответствует двум заражённым узлам, рождённым в периоде 2, из которых один размножается сразу после одного периода (то есть даёт вклад 0 при задержке $i=2$), а другой имеет задержку в два периода, порождая двух новых заражённых узлов в периоде 4; третья группа $(i=3)$ соответствует двум заражённым узлам, рождённым в периоде 1, ни один из которых не размножается при задержке в три периода; а последнее слагаемое $(i=4)$ соответствует единственному заражённому хосту в периоде 0, который не размножается при задержке в четыре периода:

$$Z(4) = (1+1+0)+(0+2)+(0+0)+0 = 4.$$

Результаты и обсуждение. Обозначим $\mathsf{E}[X_{ij}^{(n)}]$ через μ_i, т.е. среднее число частиц в $i-$поколении.

Теорема 1. (см. [1]–[2]). *Пусть* $\{Z(n),\ n\ge0\}$ *ВПЗПР и* $\mathsf{E}[X_{ij}^{(n)}] = \mu_i$, *тогда*

$$\mathsf{E}[Z(n)] = \sum_{i=1}^{n}\mu_i\mathsf{E}[Z(n-i)].$$

Хотя Теорема 1 характеризует математическое ожидание числа заражённых узлов по поколениям, она не даёт информации о том, насколько отдельная реализация $Z(n)$ может отклоняться от этого ожидаемого значения, информации, необходимой для оценки надёжности любой Монте-Карло

оценки $Z(n)$. Знание $\mathsf{Var}[Z(n)]$ позволяет построить доверительный интервал вокруг смоделированной оценки, обеспечивая тем самым количественную меру точности оценивания, дополняющую заявленный в настоящем исследовании анализ сходимости. Помимо своей методологической роли, данная величина также несёт прямую практическую интерпретацию в контексте распространения вредоносного программного обеспечения: большая дисперсия указывает на то, что число заражённых узлов в данном поколении может существенно отличаться от своего ожидаемого значения, что подразумевает большую неопределённость при прогнозировании масштаба вспышки заражения, тогда как малая дисперсия указывает на то, что динамика распространения является сравнительно предсказуемой. Следующий результат, таким образом, распространяет рекуррентную формулу математического ожидания из Теоремы 1 на соответствующее рекуррентное выражение для дисперсии $Z(n)$.

Теорема 2. *Пусть* $\{Z(n),\ n \geq 0\}$ *ВПЗПР, для которого*

$$\mathsf{E}[X_i^{(n)}] = \mu_i \quad \textit{и} \quad \mathsf{Var}[X_i^{(n)}] = \sigma_i^2$$

для каждой задержки $i \in \mathsf{N}$*, тогда*

$$\mathsf{Var}[Z(n)] = \sum_{i=1} \sigma_i^2 \mathsf{E}[Z(n-i)] + \sum_{i=1} \sum_{k=1} \mu_i \mu_k \mathsf{Cov}(Z(n-i), Z(n-k)). \tag{3}$$

Доказательство Теоремы 2. Обусловливая на

$$Z(n-1) = a(n-1),\ Z(n-2) = a(n-2), \ldots,\ Z(0) = a(0)$$

численность популяции $Z(n)$ разлагается на n независимых групп по одной для каждой задержки i, причём $i-$группа представляет собой сумму $a(n-i)$ независимых и одинаково распределённых случайных величин

$$X_{i1}^{(n)}, X_{i2}^{(n)}, \ldots, X_{i,a(n-i)}^{(n)}.$$

Следовательно,

$$\mathsf{E}[Z(n) \mid Z(n-1), Z(n-2), \ldots, Z(0)] = \sum_{i=1}^{n} a(n-i)\mu_i = \sum_{i=1}^{n} Z(n-i)\mu_i$$

и, поскольку $a(n-i)$ слагаемых внутри каждой группы независимы и имеют общую дисперсию σ_i^2, а сами n групп независимы друг от друга,

$$\mathsf{Var}[Z(n) \mid Z(n-1), Z(n-2), \ldots, Z(0)] = \sum_{i=1}^{n} a(n-i)\sigma_i^2 = \sum_{i=1}^{n} Z(n-i)\sigma_i^2.$$

Применяя формулу полной дисперсии,

$$\mathsf{Var}[Z(n)] = \mathsf{E}[\mathsf{Var}[Z(n) \mid Z(n-1), Z(n-2), \ldots, Z(0)]] +$$

$$+ \mathsf{Var}[\mathsf{E}[Z(n) \mid Z(n-1), Z(n-2), \ldots, Z(0)]] =$$

$$= \mathsf{E}\left[\sum_{i=1}^{n} \sigma_i^2 Z(n-i)\right] + \mathsf{Var}\left[\sum_{i=1}^{n} \mu_i Z(n-i)\right] =$$

$$= \sum_{i=1}^{n} \sigma_i^2 \mathsf{E}[Z(n-i)] + \sum_{i=1}^{n} \sum_{k=1}^{n} \mu_i \mu_k \mathsf{Cov}(Z(n-i), Z(n-k))$$

что и завершает доказательство. Теорема доказана.

Моделирование методом Монте-Карло. Хотя Теоремы 1 и 2 предоставляют точные рекуррентные выражения для среднего значения и дисперсии $Z(n)$, эти величины сами по себе не раскрывают полного распределительного поведения численности заражённых узлов, а также не дают выборочной траектории того, как вспышка может фактически разворачиваться во времени. На практике специалиста по кибербезопасности часто интересует не только ожидаемый масштаб распространения вредоносного ПО, но и визуализация правдоподобных траекторий $Z(n)$ при

заданной модели заражения и покоя. Моделирование методом Монте-Карло напрямую отвечает этой потребности: путём многократной генерации реализаций пары «потомство–задержка» (X, M) в соответствии с совместным распределением $f_{X,M}(x,m)$ можно построить выборочные траектории $Z(n)$ без необходимости получения замкнутого аналитического выражения для его распределения.

1.	Start
2.	Input n, $f_{X,M}(x,m)$
3.	Set $i=0$, $Z(0)=1$
4.	Set $j=1$
5.	Generate (x,m)
6.	$Z(i+m)=Z(i+m)+x$
7.	$j=j+1$
8.	If $j \le Z(i)$ Yes
9.	$i=i+1$
10.	If $i<n$ Yes
11.	Output $Z(n)$
12.	End

Рисунок 2. Алгоритм Монте-Карло

В данном разделе мы представляем алгоритм Монте-Карло для моделирования ВПЗПР в его интерпретации, связанной с распространением вредоносного ПО. Каждый смоделированный заражённый узел независимо генерирует пару (x,m), представляющую число новых заражённых узлов, которые он порождает, и число периодов, в течение которых он остаётся неактивным, прежде чем это сделать, а полученное потомство накапливается в соответствующем будущем поколении. Помимо моделирования отдельных выборочных траекторий, мы затем используем многократные независимые прогоны данного алгоритма для эмпирической оценки $\mathsf{E}[Z(n)]$ и его дисперсии, а также сравниваем полученные Монте-Карло оценки с теоретическими значениями, следующими из Теорем 1 и 2, тем самым оценивая как сходимость, так и точность предлагаемого подхода к моделированию.

Предлагаемый алгоритм Монте-Карло выполняется следующим образом. Сначала в качестве входных данных задаются целевое поколение n и совместная функция вероятности $f_{X,M}(x,m)$. Затем инициализируются счётчик поколений i и начальная заражённая популяция $Z(0)$. Для каждого из $Z(i)$ заражённых узлов, присутствующих в поколении i, в соответствии с $f_{X,M}(x,m)$ генерируется пара (x,m), и x новых заражённых узлов добавляется в будущее поколение $Z(i+m)$, что отражает задержку в активации на $m-$периодов перед распространением. После того как все $Z(i)$ узлов поколения i обработаны, счётчик поколений увеличивается, и процедура повторяется до достижения поколения n, после чего возвращается смоделированное значение $Z(n)$; см. рисунок 2.

Приложение. Разработанная в настоящей статье модель ВПЗПР может быть расширена в нескольких направлениях, актуальных для практики кибербезопасности в реальных условиях. В частности, распределение задержки M может быть оценено на основе эмпирических данных из «песочниц» для анализа вредоносного ПО или исторических журналов инцидентов, что позволит калибровать модель под конкретные семейства вредоносного ПО, а не опираться на гипотетические функции вероятности. Данная модель также может быть расширена до многотипного ветвящегося процесса для отражения гетерогенных сетевых топологий, в которых узлы различных типов демонстрируют различное поведение с точки зрения заражения и покоя. Наконец, сопряжение моделирования методом Монте-Карло с телеметрией сети в реальном времени могло бы обеспечить создание систем раннего предупреждения, сигнализирующих об аномальном росте численности по поколениям до достижения вспышкой критического порога, что представляет собой практический инструмент поддержки принятия решений для служб информационной безопасности.

Заключение. В настоящей статье была осуществлена переинтерпретация модели ВПЗПР в контексте кибербезопасности, представляющая распространение вредоносного программного обеспечения как дискретный, структурированный по поколениям процесс, в котором заражённые узлы могут оставаться неактивными в течение случайного числа периодов, прежде чем активироваться и заразить новые узлы. Опираясь на теоретические основы ВПЗПР, мы получили новый результат, характеризующий дисперсию численности заражённых узлов по поколениям, дополняющий существующую формулу математического ожидания и позволяющий провести количественную оценку точности моделирования методом Монте-Карло посредством доверительных интервалов.

ЛИТЕРАТУРА:

1. Moradi, M., Fathi Vajargah, B. (2012). Monte Carlo Simulation for Branching Process with Delay at Reproduction Periods. Advances in Computer Science and its Applications (ACSA), 1(4), 256–260.

2. Fathi Vajargah, B., Moradi, M. (2010). Branching Process with Delay at Reproduction Periods. Far East Journal of Theoretical Statistics, 30(2), 145–151.

3. Kimmel, M., Axelrod, D.E. (2002). Branching Processes in Biology. Springer-Verlag, New York.

4. Jagers, P., Sagitov, S. (2008). General branching processes in discrete time as random trees. Bernoulli, 14(4), 949–962.

5. Dimov, I. (2008). Monte Carlo Methods for Applied Scientists. World Scientific.

6. Rubinstein, R.Y., Kroese, D.P. (2008). Simulation and the Monte Carlo Method. John Wiley & Sons, New Jersey.

7. Zou, C.C., Gong, W., Towsley, D. (2002). Code Red worm propagation modeling and analysis. In Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS), 138–147.

8. Zou, C.C., Gao, L., Gong, W., Towsley, D. (2005). Monitoring and early detection of Internet worms. IEEE/ACM Transactions on Networking, 13(5), 961–974.

9. Kephart, J.O., White, S.R. (1991). Directed-graph epidemiological models of computer viruses. In Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy, 343–359.

10.Zhu, X., Huang, J. (2021). Malware propagation model for cluster-based wireless sensor networks using epidemiological theory. PeerJ Computer Science, 7, e728.

11.Benomar, Z., Ghribi, C., Cali, E., Hinsen, A., Jahnel, B. (2022). Agent-based modeling and simulation for malware spreading in D2D networks. arXiv:2201.12230.