МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

DOI: 10.25586/RNU.V9187.19.03.P.003 УДК 004.94

В.А. Минаев, М.П. Сычев, Е.В. Вайц, А.Э. Киракосян

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ЭПИДЕМИЙ КОМПЬЮТЕРНЫХ ВИРУСОВ

Применен метод системной динамики для моделирования эпидемических процессов распространения в сетях компьютерных вирусов. Описаны процессы распространения вирусов на основе SEIR- и PSIDR-моделей и выполнена их реализация в программной среде Anylogic. Проведены имитационные эксперименты, позволяющие исследовать динамику числа уязвимых, инфицированных, латентных и вылеченных хостов сети, а также определять оптимальные значения параметров моделей при заданных ограничениях на характеристики распространения вирусов. Применение имитационной модели дало новые возможности для исследования вирусных эпидемий в компьютерных сетях, решения задач управления эпидемиями, прогнозирования их течения, определения оптимальных параметров противодействия. Сделан вывод, что дальнейшим развитием работы является построение комплекса моделей, учитывающих другие процессы, связанные с распространением вирусов, и проведение дополнительной серии имитационных экспериментов с различными комбинациями факторов, включая специфические факторы конкретных критических инфраструктур.

Ключевые слова: системно-динамическая модель, информационная безопасность, имитационное моделирование, вирусная эпидемия, имитационный эксперимент.

V.A. Minaev, M.P. Sychev, E.V. Vaits, A.E. Kirakosyan

SIMULATION MODELLING OF COMPUTER VIRUSES EPIDEMICS

The method of system dynamics for simulation of computer viruses epidemic processes in networks is applied in the article. Process of computer viruses spreading in network based on SEIR and PSIDR models are described and implemented in the Anylogic software platform. Simulation experiments have been carried out to study the dynamics of number of vulnerable, infected, latent and "cured" hosts of the network, as well as to determine the optimal values of the model parameters under the given restrictions on the characteristics of viruses spreading. The use of the simulation model has given new opportunities for the study of viral epidemics in computer networks, to solve the problems of epidemic management, forecasting their course, determining optimal parameters of counteraction. It is concluded that the further development of the investigation is the construction of a complex of models that take into account other processes associated with the spread of viruses, and an additional series of simulation experiments with various combinations of factors, including specific factors of specific critical of infrastructures.

Keywords: system-dynamic model, information security, simulation modeling, virus epidemic, simulation experiment.

Введение

Вирусные атаки представляют серьезную угрозу для всех пользователей компьютерных сетей, в том числе защищенных. Их успешные реализации могут привести к значительному ущербу и повлечь катастрофические последствия в различных критически важных секторах государства.

Своевременное обнаружение вредоносных программ и оперативное устранение последствий их деятельности, как показали события двухлетней давности, связанные с нападением в мае 2017 г. компьютерного вируса WannaCry на информационные ресурсы Российской Федерации, играет огромную роль в обеспечении информационной безопасности и устойчивости государственных служб и подразделений, бизнес-структур различного уровня и направленности.

Очевидно, что сетевые атаки на этом не закончились. В конце июня 2019 г. получил распространение очередной опасный вирус-шифровальщик Troldesh (Shade). Впереди – новые испытания для информационных систем, в том числе относящихся к критической инфраструктуре России. Поэтому исследование динамики распространения компьютерных сетевых вирусов является весьма актуальной задачей на современном этапе развития информационной инфраструктуры страны [1; 2].

В научных работах по данной проблематике дано описание целого ряда математических моделей распространения компьютерных вирусов в сетях [2; 3; 4; 5]. В то же время пока недостаточное внимание уделяется имитационному моделированию, дающему исследователям широкий спектр возможностей для решения задач анализа, оценки и прогнозирования процессов заражения компьютерных сетей вирусами [6]. Больше всего для имитационного моделирования указанных процессов подходят системно-динамические модели.

Системно-динамическое моделирование – направление в изучении сложных систем, исследующее их поведение во времени в зависимости от структуры элементов системы и взаимодействия между ними. Метод предложил Дж. Форрестер в конце 1950-х гг. Моделируемые процессы отображаются в виде некоторой структуры, состоящей из накопителей – уровней, соединенных взаимосвязанными потоками, которые, «перетекая», изменяют значение уровней [7].

Созданные до сегодняшнего дня модели динамики распространения компьютерных вирусов по сети, как правило, основываются на моделях эпидемических процессов [8; 9]. Самыми простыми моделями этого типа являются SI-модель (Susceptible – Infected model) и SIR-модель (Susceptible – Infected – Removed model).

В данной статье рассмотрим их усложненные модификации: SEIR-модель (Susceptible – Exposed – Infected – Removed model) и PSIDR-модель (Progressive Susceptible – Infected – Detected – Removed model).

Описание вирусных эпидемий на основе SEIR-модели

В SEIR-модели учитывается возможность того, что вирус может иметь некий «латентный период», во время которого он не наносит вреда инфицированному узлу. Обычно вирус заражает уязвимый узел (S) до входа в свою латентную стадию. В течение латентного периода (Ex, *Exposed*) узел считается зараженным, но не распространяет вирус. Через некоторое время он становится способным к заражению других хостов (I) и далее превращается в «излеченный» (R) (рис. 1).

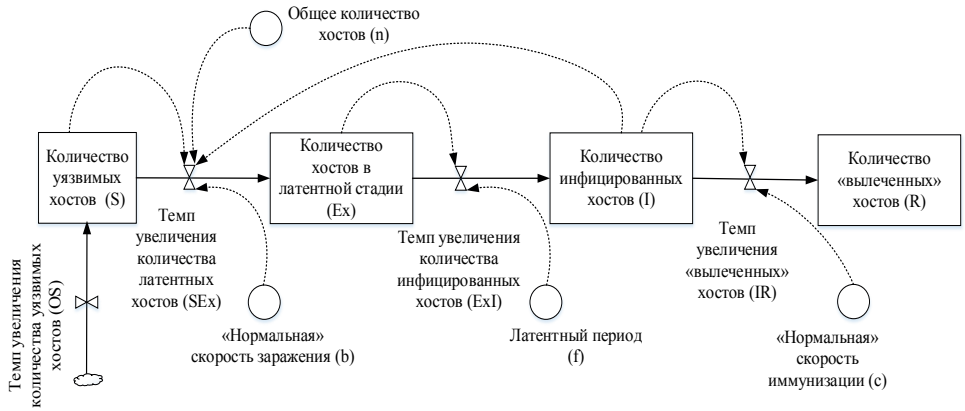

Рис. 1. Системно-динамическая SEIR-модель распространения компьютерных вирусов по сети

SEIR-модель описывается следующей системой уравнений:

$$\begin{cases} dS/dt = OS(t) - SEx(t), \\ dEx/dt = SEx(t) - ExI(t), \\ dI/dt = ExI(t) - IR(t), \\ dR/dt = IR(t), \\ SEx(t) = \left[bS(t)I(t) \right] / n, \\ ExI(t) = Ex(t)/f, \\ IR(t) = cI(t). \end{cases} \quad (1)$$

Расшифровка обозначений, используемых в SEIR-модели, приведена в таблице 1.

Таблица 1

Условные обозначения, используемые в SEIR-модели

№ п/п	Условное обозначение элемента	Название элемента (единица измерения)
1	*S*	Количество уязвимых хостов (шт.)
2	*Ex*	Количество инфицированных узлов, находящихся в латентной стадии (шт.)
3	*I*	Количество инфицированных хостов (шт.)
4	*R*	Количество «излеченных» хостов (шт.)
5	*n*	Общее количество хостов в сети (шт.)
6	*OS*	Темп увеличения новых уязвимых хостов (шт./ч)
7	*SEx*	Темп увеличения латентных хостов (шт./ч)
8	*ExI*	Темп увеличения инфицированных хостов (шт./ч)
9	*IR*	Темп увеличения «излеченных» хостов (шт./ч)
10	*b*	«Нормальная» скорость заражения (доля/ч)
11	*f*	Латентный период (ч)
12	*c*	«Нормальная» скорость «иммунизации» (доля/ч)

Понятие «нормальной» скорости, введенное Дж. Форрестером [7], представляет отношение числа зараженных или излеченных хостов в день к общему количеству уязвимых хостов.

Реализована модель распространения компьютерных вирусов по сети на базе SEIR-модели в программной среде Anylogic. Общий вид интерфейса модели представлен на рисунке 2.

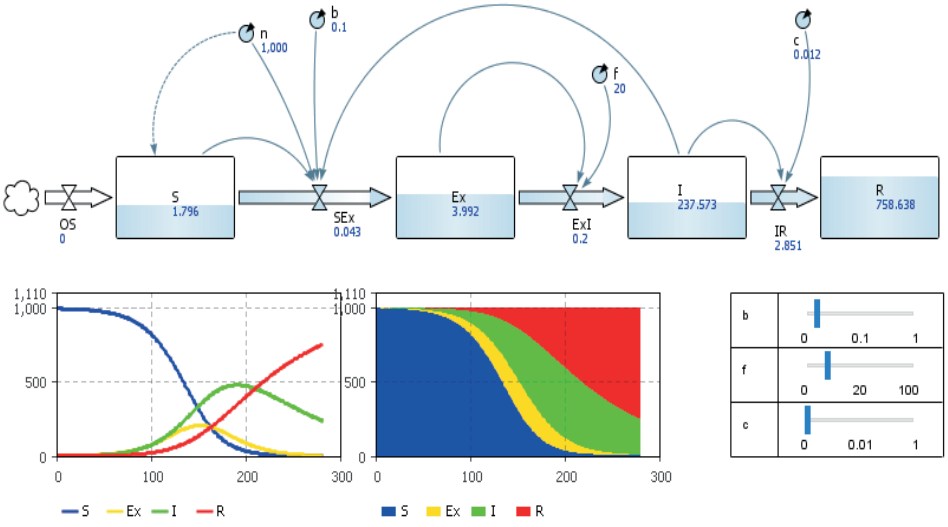

Рис. 2. Общий вид интерфейса SEIR-модели распространения компьютерных сетевых вирусов

С моделью проведен имитационный эксперимент, в котором определялось минимальное значение скорости иммунизации при заданном ограничении на максимальное количество инфицированных хостов сети.

Математическая постановка такой задачи выглядит следующим образом:

$$\begin{cases} c \to \min, \\ I \le I_{\max}. \end{cases} \qquad (2)$$

Диапазон возможных значений скорости иммунизации примем следующим: $c \in \{0{,}001;\ 0{,}03\}$, шаг имитации при проведении эксперимента – 0,001.

Начальные значения других параметров модели определены следующим образом: $S(0) = n = 1\ 000$; $I(0) = 2$; $R(0) = Ex(0) = 0$; $b = 0{,}1$; $f = 20$.

Оптимизационный эксперимент проведен с использованием встроенного в программу Anylogic специального алгоритма OptQuest, в котором используются как точные методы математической оптимизации, так и нейронные сети и эвристические подходы к поиску решений.

Приведем зависимости значений оптимизируемого параметра (c) от номера итерации при $I_{\max} = 47$ и $I_{\max} = 57$ (рис. 3).

На графиках нижней ступенчатой функцией отображено лучшее недопустимое значение, т.е. значение, полученное без учета ограничений, наложенных на оптимизируемую модель, а верхней ступенчатой функцией – лучшее допустимое. Можно легко заметить, что с увеличением количества итераций значение оптимизируемого параметра стремится к наилучшему значению целевой функции.

Результаты эксперимента: при ограничении количества инфицированных хостов I_{max} = = 47 минимальная скорость иммунизации должна составлять 0,02, а при ограничении количества инфицированных хостов I_{max} = 57 должна составлять 0,017.

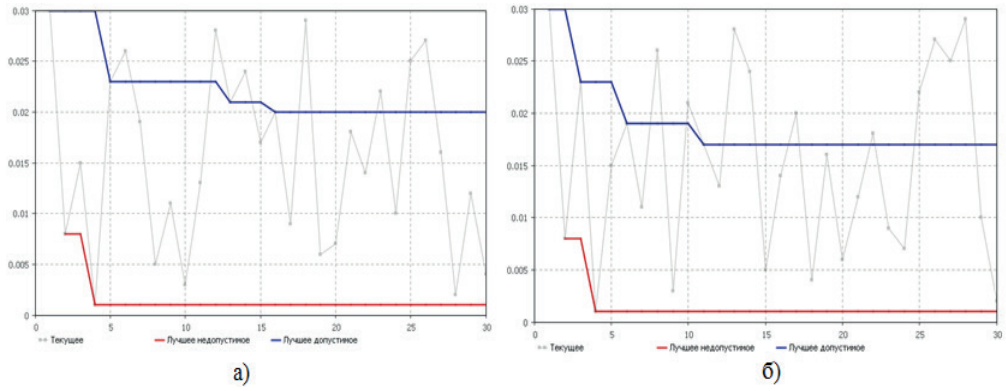

Рис. 3. Зависимости значений оптимизируемого параметра *c* от номера итерации:
а – при I_{max} = 47; *б* – при I_{max} = 57

Описание вирусных эпидемий на основе PSIDR -модели

В PSIDR-модели предполагается, что эпидемические события разделены на два периода:

- *Начальный период.* Изначально вирус инфицирует один хост в сети. После этого в течение определенного времени вирус распространяется по сети, будучи не замеченным ее пользователями. Этот период характеризуется скоростью заражения *b* без попыток излечения. В модели начальный период обозначим как τ.
- *Период реакции на вирус.* Через период времени τ вирус обнаруживается. Осуществляется выделение его сигнатур и внесение их в базы антивирусного программного обеспечения. Неинфицированные узлы становятся невосприимчивыми к данному вирусу, а инфицированные хосты «излечиваются» по мере обновления антивирусных баз.

Таким образом, PSIDR-модель предполагает, что течение эпидемии можно разбить на два периода: вначале система может находиться в двух состояниях $S \to R$, а по истечении времени τ система переходит в состояния $S \to I \to D \to R$ с возможностью и прямого перехода между состояниями $S \to R$.

Построим схему распространения компьютерного вируса по сети на основе PSIDR-модели (рис. 4). Приведем расшифровку обозначений, используемых в PSIDR-модели (табл. 2).

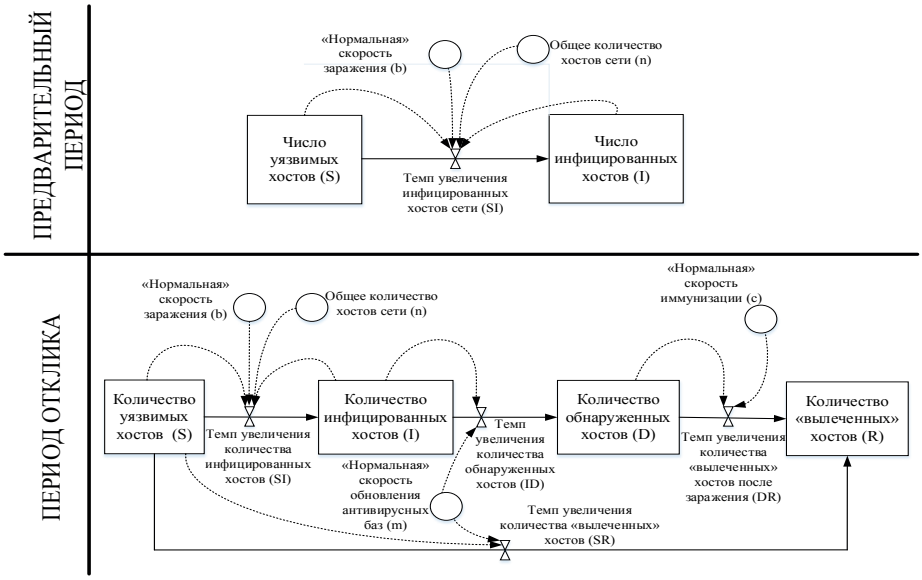

Рис. 4. Системно-динамическая PSIDR-модель распространения компьютерных вирусов по сети

Таблица 2

Условные обозначения, используемые в PSIDR-модели

№ п/п	Условное обозначение элемента	Название элемента (единица измерения)
1	*n*	Общее количество хостов сети (шт.)
2	*S*	Количество уязвимых хостов (шт.)
3	*I*	Количество инфицированных хостов (шт.)
4	*D*	Количество обнаруженных хостов (шт.)
5	*R*	Количество «вылеченных» хостов (шт.)
6	*SI*	Темп увеличения количества инфицированных хостов (шт./ч)
7	*ID*	Темп увеличения количества обнаруженных хостов (шт./ч)
8	*DR*	Темп увеличения количества «вылеченных» хостов после заражения (шт./ч)
9	*SR*	Темп увеличения количества «вылеченных» хостов (шт./ч)
10	*b*	«Нормальная» скорость заражения (доля/ч)
11	*c*	«Нормальная» скорость иммунизации (доля/ч)
12	*m*	«Нормальная» скорость обновления антивирусных баз (доля/ч)

В начальный период модель описывается следующей системой уравнений:

$$\begin{cases} dS/dt = -SI(t), \\ dI/dt = SI(t), \\ SI(t) = \left[bS(t)I(t)\right]/n. \end{cases} \quad (3)$$

В период реакции модель начинает описываться следующей системой уравнений:

$$\begin{cases} dS/dt = -SI(t) - SR(t), \\ dI/dt = SI(t) - ID(t), \\ dD/dt = ID(t) - DR(t), \\ dR/dt = DR(t) + SR(t), \\ SI(t) = \left[bS(t)I(t) \right] / n, \\ ID(t) = mI(t), \\ DR(t) = cD(t), \\ SR(t) = mS(t). \end{cases} \quad (4)$$

Схема распространения вируса по сети на основе PSIDR-модели реализована в программе Anylogic. Общий вид интерфейса модели в начальный период представлен на рисунке 5, в период реакции на вирус – на рисунке 6.

Проведено два имитационных эксперимента, в которых исследуется влияние скорости заражения, скорости обновления антивирусных баз и скорости иммунизации сети на количество уязвимых, инфицированных, обнаруженных и «вылеченных» хостов сети.

Задачей данного эксперимента является исследование поведения модели в указанных двух периодах: начальный период и период реакции (табл. 3).

Таблица 3

Значения параметров и исследуемые переменные модели в экспериментах

Номер эксперимента	Значение параметров модели		Значение переменных модели			
	m	***c***	***S***	***I***	***D***	***R***
1	0	0	$S_1(t)$	$I_1(t)$	$D_1(t)$	$R_1(t)$
2	0,04	03	$S_2(t)$	$I_2(t)$	$D_2(t)$	$R_2(t)$

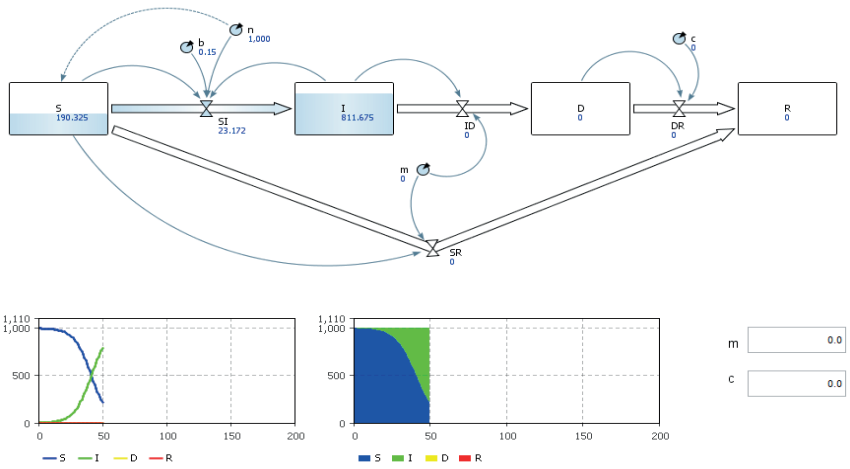

Рис. 5. Общий вид интерфейса PSIDR-модели распространения вируса по сети в начальный период

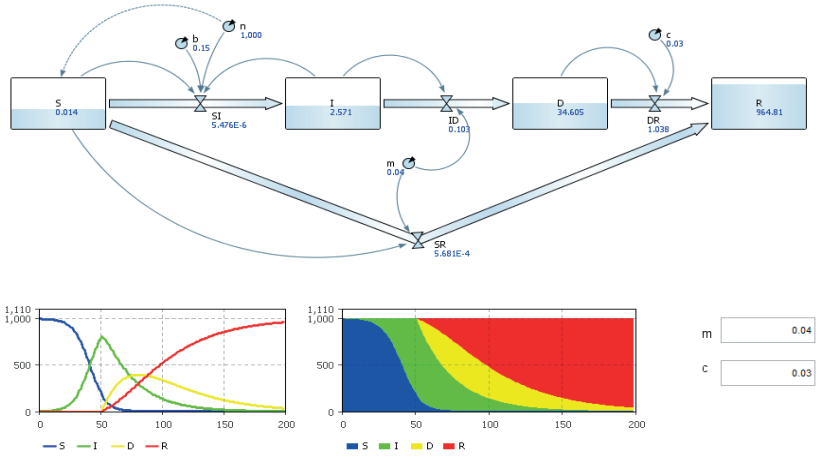

Рис. 6. Общий вид интерфейса PSIDR-модели распространения вируса по сети в период реакции

Начальные значения переменных и параметра b модели определим следующим образом: $S(0) = n = 1000$; $I(0) = 2$; $R(0) = 0$; $D(0) = 0$; $b = 0{,}15$.

Приведенные на рисунке 7 кривые отражают динамику уязвимых, инфицированных, обнаруженных и «излеченных» хостов сети во время двух периодов работы модели.

Первый – начальный (предварительный) – период (при $t < \tau$) отражает только динамику уязвимых и инфицированных хостов сети, так как параметры m – «нормальная» скорость обновления антивирусных баз и c – «нормальная» скорость «иммунизации» в это время равны нулю.

Второй период – реакции (отклика) (при $t \geq \tau$) отражает уже динамику всех рассматриваемых состояний системы. В экспериментах длительность начального периода $\tau = 50$ ч.

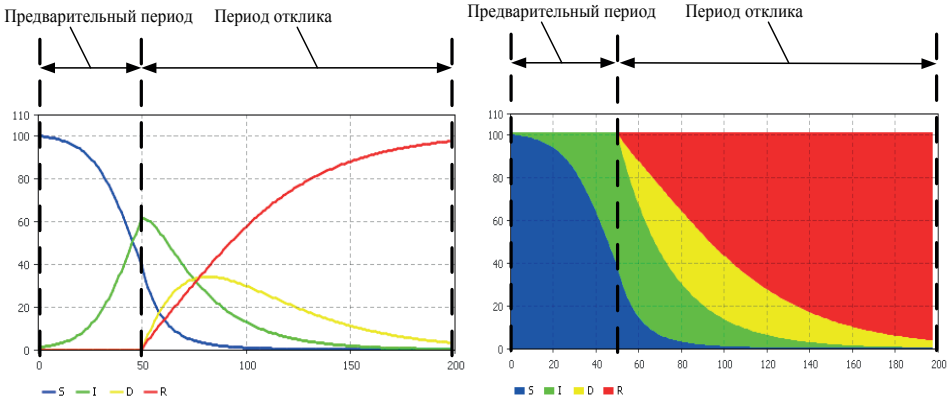

Рис. 7. Динамика состояний PSIDR-модели распространения вируса по сети в двух периодах (*б* – диаграмма с накоплением)

Таким образом, эксперименты наглядно демонстрируют качественные различия работы модели на различных временных отрезках (начальном периоде и периоде реакции).

Выводы

1. Применение имитационного моделирования предоставляет новые возможности для исследования вирусных эпидемий в компьютерных сетях, а именно возможность решать задачи управления эпидемиями, прогнозирования их течения, определения оптимальных параметров противодействия и др.

2. Системно-динамические модели распространения компьютерных вирусов по сети, основывающиеся на эпидемических SEIR- и PSIDR-моделях, позволяют исследовать динамику «заражения» сети и выявлять степень влияния наиболее критичных факторов. Реализация построенных моделей в программной среде Anylogic дает возможность наглядно отображать эпидемии компьютерных вирусов при различных значениях параметров модели.

3. Проведенные имитационные эксперименты позволяют прогнозировать динамику числа уязвимых, инфицированных, латентных, обнаруженных и «излеченных» хостов сети в зависимости от различных значений параметров моделей, а также определять оптимальные значения параметров моделей при заданных ограничениях на характеристики распространения вирусов.

4. Дальнейшим развитием работы является построение комплекса моделей, учитывающих другие процессы, связанные с распространением вирусов [10; 11; 12], и проведение дополнительной серии имитационных экспериментов с различными комбинациями учтенных факторов, включая специфические факторы конкретных критических инфраструктур.

Литература

1. О безопасности критической информационной инфраструктуры Российской Федерации: федеральный закон от 26.07.2017 № 187-ФЗ. Доступ из справ.-правовой системы «КонсультантПлюс».

2. *Борзунов К.К. и др.* Информационные технологии в органах внутренних дел: монография / под ред. А.С. Овчинского. М.: Московский университет МВД России, 2010. 345 с.

3. *Котенко И.В., Воронцов В.В.* Аналитические модели распространения сетевых червей // Труды СПИИРАН. СПб.: Наука, 2007. С. 208–224.

4. *Захарченко А.* Черводинамика: причины и следствия // Защита информации. Конфидент. 2004. № 2. С. 50–55.

5. *Семыкина Н.А., Шавыкина И.В.* Математическая модель защиты компьютерной сети от вирусов // Программные продукты и системы. 2016. Т. 29, № 4. С. 125–128.

6. *Семенов С.Г., Давыдов В.В.* Математическая модель распространения компьютерных вирусов в гетерогенных компьютерных сетях автоматизированных систем управления технологическим процессом // НТУ «ХПИ». Серия: Информатика и моделирование. 2012. № 38. С. 163–171.

7. *Аверенков В.И., Федоров В.П., Хейфец М.Л.* Основы математического моделирования технических систем. Брянск: Изд-во БГТУ, 2004. 271 с.

8. *Форрестер Д.* Основы кибернетики предприятия (индустриальная динамика). М.: Прогресс, 1971. 340 с.

9. *Kephart J.O., White S.R.* Directed-Graph Epidemiological Models of Computer Viruses // Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy. Oakland, 1991. P. 343–359.
10. *Минаев В.А., Сычев М.П., Вайц Е.В., Грачева Ю.В.* Математическая модель «хищник-жертва» в системе информационной безопасности // Информация и безопасность. 2016. Т. 19, № 3 (4). С. 397–400.
11. *Чекалкин А.А., Скрыль С.В., Минаев В.А.* Комплексный технический контроль эффективности мер безопасности систем управления в органах внутренних дел: учебное пособие для высших учебных заведений МВД России. Ч. 2: Практические аспекты технической разведки и комплексного технического контроля. М.: Горячая линия – Телеком, 2006. 205 с.
12. *Кулаков В.Г. и др.* Защита информации в телекоммуникационных системах: учебник для высших учебных заведений МВД России. Воронеж: Воронежский институт МВД России, 2002. 300 с.

Literatura

1. O bezopasnosti kriticheskoj informacionnoj infrastruktury Rossijskoj Federacii: federal'nyj zakon ot 26.07.2017 № 187-FZ. Dostup iz sprav.-pravovoj sistemy "Konsul'tantPlyus".
2. *Borzunov K.K. i dr.* Informacionnye tekhnologii v organah vnutrennih del: monografiya / pod red. A.S. Ovchinskogo. M.: Moskovskij universitet MVD Rossii, 2010. 345 s.
3. *Kotenko I.V., Voroncov V.V.* Analiticheskie modeli rasprostraneniya setevyh chervej // Trudy SPIIRAN. SPb.: Nauka, 2007. S. 208–224.
4. *Zaharchenko A.* Chervodinamika: prichiny i sledstviya // Zashchita informacii. Konfident. – 2004. № 2. S. 50–55.
5. *Semykina N.A., Shavykina I.V.* Matematicheskaya model' zashchity komp'yuternoj seti ot virusov // Programmnye produkty i sistemy. 2016. T. 29, № 4. S. 125–128.
6. *Semenov S.G., Davydov V.V.* Matematicheskaya model' rasprostraneniya komp'yuternyh virusov v geterogennyh komp'yuternyh setyah avtomatizirovannyh sistem upravleniya tekhnologicheskim processom // NTU "HPI". Seria: Informatika i modelirovanie. 2012. № 38. S. 163–171.
7. *Averenkov V.I., Fedorov V.P., Hejfec M.L.* Osnovy matematicheskogo modelirovaniya tekhnicheskih sistem. Bryansk: Izd-vo BGTU, 2004. 271 s.
8. *Forrester D.* Osnovy kibernetiki predpriyatiya (industrial'naya dinamika). M.: Progress, 1971. 340 s.
9. *Kephart J.O., White S.R.* Directed-Graph Epidemiological Models of Computer Viruses // Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy. Oakland, 1991. P. 343–359.
10. *Minaev V.A., Sychev M.P., Vajc E.V., Grachyova Yu.V.* Matematicheskaya model' "hishchnik-zhertva" v sisteme informacionnoj bezopasnosti // Informaciya i bezopasnost'. 2016. T. 19, № 3 (4). S. 397–400.
11. *Chekalkin A.A., Skryl' S.V., Minaev V.A.* Kompleksnyj tekhnicheskij kontrol' effektivnosti mer bezopasnosti sistem upravleniya v organah vnutrennih del: uchebnoe posobie dlya vysshih uchebnyh zavedenij MVD Rossii. Ch. 2: Prakticheskie aspekty tekhnicheskoj razvedki i kompleksnogo tekhnicheskogo kontrolya. M.: Goryachaya liniya – Telekom, 2006. 205 s.
12. *Kulakov V.G. i dr.* Zashchita informacii v telekommunikacionnyh sistemah: uchebnik dlya vysshih uchebnyh zavedenij MVD Rossii. Voronezh: Voronezhskij institut MVD Rossii, 2002. 300 p.