

MỘT SỐ KỸ THUẬT XÂY DỰNG MÔ HÌNH NGẪU NHIÊN

Trần Hải Anh, Đỗ Xuân Chợ, Nguyễn Đức Việt

Học Viện Công Nghệ Bưu Chính Viễn Thông

Tóm tắt: Trong những năm gần đây, sự tấn công của mã độc đã trở thành một thách thức lớn và đe dọa tới nhiều tổ chức, doanh nghiệp cho đến các cá nhân. Và để giải quyết bài toán này, những nhà quản trị mạng đã áp dụng các mô hình toán học được sử dụng rộng rãi trong dịch tễ học để theo dõi xu hướng và cường độ tấn công của mã độc, điều này có thể cung cấp một số hỗ trợ về mặt lý thuyết cho các quyết định dự đoán và kiểm soát mã độc trong tương lai. Trên thực tế đã có rất nhiều mô hình được đề xuất, tuy nhiên nhiều nhà nghiên cứu có xu hướng sử dụng các mô hình bất định vì sự đơn giản của nó nhưng trên thực tế thì không phải lúc nào loại mô hình này cũng phù hợp. Các sự kiện ngẫu nhiên khó đoán trước thường xuất hiện, điều này làm giảm đi giá trị mà mô hình bất định đem lại. Để có thể giải quyết được vấn đề này, một số nhà nghiên cứu đã cho ra các phương pháp ngẫu nhiên nhằm tăng tính thực tế của mô hình. Chính vì vậy, trong bài báo này, chúng tôi muốn trình bày một số quá trình ngẫu nhiên thường được áp dụng trong các mô hình ngẫu nhiên hiện nay để ứng dụng vào lĩnh vực an ninh mạng.

Từ khóa: mô hình ngẫu, quá trình ngẫu nhiên, xích Markov.

I. GIỚI THIỆU

Những năm trở lại đây, việc áp dụng các mô hình toán học nhằm dự đoán xu thế lây lan của virus trong dịch tễ học cũng như mã độc trong an toàn thông tin đã nhận được nhiều sự quan tâm từ các nhà nghiên cứu. Sự tương đồng giữa virus và mã độc đã mở đường cho hàng loạt các nghiên cứu [11,14,15], chúng góp phần cải thiện khả năng dự báo nhằm cung cấp thêm cơ sở để các nhà quản trị mạng đưa ra những quyết định chính xác và nhanh chóng trước sự tấn công trên không gian mạng ngày càng dày đặc và phức tạp. Sự đơn giản của các mô hình toán học giúp chúng ta tập trung nghiên cứu vào những đặc tính quan trọng nhất thay vì phân tán trên nhiều đặc điểm khác nhau với yêu cầu tốn kém về mặt tài nguyên và nhân lực để có thể đưa ra một dự đoán nào đó. Vậy mô hình toán học được ứng dụng trong dịch tễ học là gì?!

Mô hình toán học này còn được gọi là mô hình dịch bệnh hay mô hình ngẫu xuất hiện lần đầu tiên vào năm 1911

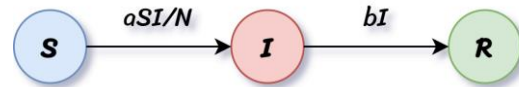
nhằm phục vụ cho mục đích nghiên cứu về bệnh sốt rét được phát triển bởi Ross [1,2,3], nó cũng là cơ sở cho vô số mô hình mô tả vector truyền nhiễm sau này. Dưới đây là mô hình với một hệ gồm hai phương trình vi phân:

$$\begin{cases} \frac{dI}{dt} = bp'i \frac{N-I}{N} - aI, \\ \frac{di}{dt} = bp(n-i) \frac{I}{N} - mi. \end{cases} \quad (1)$$

với N : tổng dân số trong vùng; $I(t)$: tổng dân số bị nhiễm bệnh tại thời điểm t ; n : tổng số muỗi; $i(t)$: tổng số muỗi bị nhiễm mầm bệnh; b : tần suất muỗi đốt; p : xác suất mầm bệnh truyền từ người qua muỗi; p' : xác suất mầm bệnh truyền từ muỗi sang người; a : tỷ lệ dân số phục hồi; m : tỷ lệ muỗi chết.

Tiếp nối thành công từ mô hình của Ross, các nhà nghiên cứu đã sửa đổi mô hình này để giải thích các yếu tố mà họ muốn định nghĩa, một số như mô hình Macdonald, mô hình Anderson-May ..v.v. Tuy nhiên, có một số mô hình tiêu biểu được sử dụng trong nhiều thập kỷ qua như mô hình SI, mô hình SIS và đặc biệt phải kể đến là mô hình cổ điển SIR được Kermack và McKendrick [4] phát biểu năm 1927. Đây là mô hình được sử dụng để mô tả diễn biến lây lan của một bệnh truyền nhiễm hay một loại mã độc trong một khoảng thời gian xác định. Nó được xây dựng từ ba loại trạng thái: S (Susceptible), I (Infectious) và R (Recovered), mỗi một loại đều mô tả một tỷ lệ hoặc số lượng người tại từng thời điểm.

• Mô hình SIR:



Hình 1: Mô hình SIR cổ điển

Hệ phương trình vi phân của mô hình SIR cổ điển như sau:

$$\begin{cases} \frac{dS}{dt} = -\frac{aSI}{N}, \\ \frac{dI}{dt} = \frac{aSI}{N} - bI, \\ \frac{dR}{dt} = bI. \end{cases} \quad (2)$$

với a : tỷ lệ truyền nhiễm và b : tỷ lệ phục hồi. Vì hệ (2) có số lượng không đổi $N(t) = S(t) + I(t) + R(t)$ nên hệ có thể giảm xuống còn hai phương trình vi phân.

Hệ số lây nhiễm cơ bản:

Tác giả liên hệ: Trần Hải Anh,

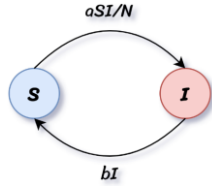
Email: haianh121293@gmail.com

Đến tòa soạn: 06/2024, chỉnh sửa: 07/2024, chấp nhận đăng:

08/2024.

$$R_0 = \frac{a}{b} \quad (3)$$

- Mô hình SIS:



Hình 2: Mô hình SIS

Hệ phương trình vi phân mô tả hoạt động của mô hình SIS được minh họa trong hình 2 như sau:

$$\begin{cases} \frac{dS}{dt} = -\frac{aSI}{N} + bI, \\ \frac{dI}{dt} = \frac{aSI}{N} - bI. \end{cases} \quad (4)$$

với $S + I = N$, a : tỷ lệ truyền nhiễm và b : tỷ lệ phục hồi, tương tự như mô hình SIR với số lượng N phần tử không đổi theo thời gian t , ta có thể đưa về một phương trình vi phân duy nhất.

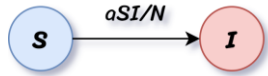
Hệ số lây nhiễm cơ bản:

$$R_0 = \frac{a}{b} \quad (5)$$

Trạng thái cân bằng dịch:

$$S = \frac{bN}{a}, \quad I = \frac{(a-b)N}{a} = \left(1 - \frac{b}{a}\right)N \quad (6)$$

- Mô hình SI:



Hình 3: Mô hình SI

Hệ phương trình vi phân biểu diễn hoạt động của mô hình SI được minh họa như sau:

$$\begin{cases} \frac{dS}{dt} = -\frac{aSI}{N}, \\ \frac{dI}{dt} = \frac{aSI}{N}. \end{cases} \quad (7)$$

với $S + I = N$, a : tỷ lệ truyền nhiễm, tương tự như các mô hình trên mô hình SI cũng có thể đưa về một phương trình vi phân duy nhất.

Để thấy tất cả những mô hình ban đầu như (1), (2), (4) và (7) đều là các mô hình tất định, chúng thích hợp khi sử dụng với số lượng phần tử lớn như nhân khẩu học trong một thành phố. Nhưng khi quần thể nghiên cứu bị hạn chế trong một phạm vi nhỏ như số lượng máy tính của một công ty thì các mô hình ngẫu nhiên là một giải pháp phù hợp hơn so với các mô hình tất định [5,6,7,12].

Mô hình tất định (Deterministic model) là mô hình toán học với cách tiếp cận đơn giản với các tham số đã được xác định. Mô hình này luôn cho ra cùng một kết quả cho cùng một đầu vào [12].

Mô hình ngẫu nhiên (Stochastic model) là mô hình áp dụng quá trình ngẫu nhiên nhằm mô tả vòng đời của một quần thể được nghiên cứu có phụ thuộc cả vào các quá trình riêng biệt bên trong lẫn các yếu tố môi trường bên ngoài, vì vậy mô hình này thích hợp cho các sự kiện phức tạp với

kết quả thu được là không chắc chắn hoặc thay đổi như đánh giá rủi ro hay dự đoán thị trường chứng khoán [12].

Bảng 1: So sánh mô hình tất định và ngẫu nhiên

Mô hình tất định	Mô hình ngẫu nhiên
Phương trình vi phân	Dựa trên quá trình phân nhánh, xích Markov, ...
Xác định mã độc có bị tiêu diệt hay không?	Sự bùng phát khác nhau
Không phù hợp với dân số nhỏ/không thực tế	Cần tính toán nhiều và yêu cầu dữ liệu chi tiết hơn

Về mặt bản chất có thể coi là một chuỗi các xác suất. Tùy từng mô hình mà thực hiện xây dựng quá trình ngẫu nhiên theo nhiều cách khác nhau. Cụ thể hơn một mô hình ngẫu nhiên có một hoặc nhiều phần tử hoặc thành phần ngẫu nhiên và không giống như các mô hình tất định với một đầu vào chỉ duy nhất một đầu ra thì với mô hình ngẫu nhiên, một đầu vào có thể dẫn đến nhiều đầu ra khác nhau cho mỗi lần chạy, do thành phần ngẫu nhiên của quá trình đã được mô hình hóa cho các yếu tố ngẫu nhiên ảnh hưởng đến thay đổi số lượng phần tử trong quần thể. Các hệ ngẫu nhiên thường được mô tả bằng việc sử dụng một số loại quá trình ngẫu nhiên ví dụ như Ehrenfest, Poisson, Bernoulli-Laplace, xích Markov thời gian liên tục (CTMC), xích Markov thời gian rời rạc (DTMC).

Trong bài báo này chúng tôi sẽ trình bày lý thuyết một cách ngắn gọn về quá trình xích Markov thời gian liên tục và xích Markov thời gian rời rạc trong mục II. Tiếp theo là quá trình mô phỏng số cho các mô hình ngẫu nhiên dựa trên quá trình ngẫu nhiên sẽ được trình bày tại mục III.

II. MỘT SỐ QUÁ TRÌNH NGẪU NHIÊN.

A. Xích Markov thời gian rời rạc

Chuỗi Markov là một tập các biến ngẫu nhiên $X = (X_n)_{n \in T}$ với T là tập chỉ số thời gian đếm được thuộc tập số nguyên dương $\mathbb{Z}_+ := \{0, 1, \dots\}$ hay tập số tự nhiên $\mathbb{N}$. Với mỗi X_i nhận các giá trị trong $(X, \mathcal{B}(X))$ với $\mathcal{B}(X)$ là một σ -đại số của tất cả các tập con của X ($(X, \mathcal{B}(X))$ được gọi là một không gian đo được hay không gian Borel). Sau cùng với mỗi $i \in X$, ta có độ đo $\mathbb{P}_i$ sao cho xác suất của biến cố X đều có thể đo được trên $\mathcal{B}(X)$. Không gian xác suất $(X, \mathcal{B}(X), \mathbb{P})$ còn được gọi là không gian độ đo xác định một quá trình ngẫu nhiên [8].

Định nghĩa 1: Quá trình $X = (X_0, X_1, \dots)$ nhận các giá trị trong $(X, \mathcal{B}(X), \mathbb{P})$ là một chuỗi Markov nếu với mọi n và chuỗi các trạng thái bất kỳ $\{i_0, i_1, i_2, \dots, i_{n-1}, i, j\}$

$$\begin{aligned} \mathbb{P}_\mu\{X_0 = i_0, X_1 = i_1, \dots, X_n = i, X_{n+1} = j\} \\ = \mu(i_0)\mathbb{P}_{i_0}(X_1 = i_1) \dots \mathbb{P}_{i_{n-1}}(X_n = i)\mathbb{P}_i(X_{n+1} = j) \end{aligned} \quad (8)$$

với μ là phân phối ban đầu của chuỗi. Và quá trình X là một chuỗi Markov nếu tại mọi thời điểm t với mọi giá trị $i_0, i_1, i_2, \dots, i_{n-1}, i, j$ cho trước với luật xác suất cho tương lai độc lập với quá khứ và chỉ phụ thuộc vào hiện tại được biểu diễn bằng công thức xác suất có điều kiện của quá trình X mô tả xác suất chuyển từ thời điểm n sang $n + 1$ như sau:

$$\begin{aligned} \mathbb{P}\{X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i\} \\ = P\{X_{n+1} = j | X_n = i\} \end{aligned} \quad (9)$$

Công thức (8) còn được gọi là tính Markov, qua đó công thức (9) có thể được viết lại như sau:

$$\begin{aligned} & \mathbb{P}_\mu\{X_0 = i_0, X_1 = i_1, \dots, X_n = i, X_{n+1} = j\} \\ &= \mu(i_0)P\{X_1 = i_1 | X_0 = i_0\} \dots P\{X_{n+1} = j | X_n = i\} \end{aligned} \quad (10)$$

Trong ứng dụng chuỗi Markov vào thực tế, người ta luôn xác định trước $\mathbb{P}_{i_0}$ và các phân bố tổng thể dựa trên xác suất chuyển một bước thông qua công thức (10). Điều này có thể thực hiện được thông qua ma trận xác suất chuyển.

Định nghĩa 2: Ma trận $\mathbf{P} = \{P_{ij}, i, j \in \mathcal{X}\}$ được gọi là ma trận xác suất chuyển hay ma trận ngẫu nhiên (stochastic matrix) nếu

$$P_{ij} \geq 0, \quad \sum_{j=0}^{\infty} P_{ij} = 1, \quad i, j \in \mathcal{X} \quad (11)$$

Có thể biểu diễn ma trận $\mathbf{P}$ dưới dạng ma trận vuông vô hạn như sau:

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots & P_{0j} & \dots \\ P_{10} & P_{11} & P_{12} & \dots & P_{1j} & \dots \\ P_{20} & P_{21} & P_{22} & \dots & P_{2j} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\ P_{i0} & P_{i1} & P_{i2} & \dots & P_{ij} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix},$$

với $\mathbf{P}(0) = \mathbf{I}$ ($\mathbf{I}$ là ma trận đơn vị).

Giả sử ta có phân phối xác suất tại trạng thái ban đầu X_0 được định nghĩa là π_0 với tập trạng thái $\mathcal{S} = \{1, 2, \dots, k\}$:

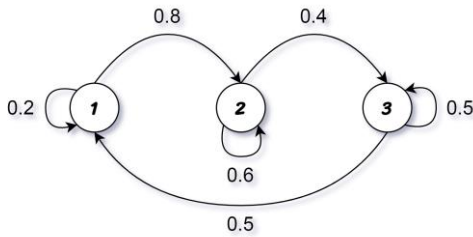
$$\pi^{(0)} = [P(X_0 = 1), P(X_0 = 2), \dots, P(X_0 = k)] \quad (12)$$

Phân bố xác suất tại thời điểm t thu được:

$$\pi^{(t+1)} = \pi^{(t)} \mathbf{P} \quad \text{với } t = 0, 1, 2, \dots \quad (13)$$

$$\pi^{(t)} = \pi^{(0)} \mathbf{P}^t \quad \text{với } t = 0, 1, 2, \dots$$

Ví dụ: Một chuỗi Markov với phân phối xác suất tại thời điểm ban đầu $\pi^{(0)} = [1, 0, 0]$.



Hình 4: Ví dụ về sơ đồ chuyển trạng thái

Từ hình 4 ta thu được ma trận xác suất chuyển sau:

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 0.2 & 0.8 & 0 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{bmatrix} \end{matrix}$$

Để xác định được phân phối dừng, ta chỉ cần giải hệ phương trình:

$$\begin{cases} 0.2\pi_1 + 0.5\pi_3 = \pi_1 \\ 0.8\pi_1 + 0.6\pi_2 = \pi_2 \\ 0.4\pi_2 + 0.5\pi_3 = \pi_3 \end{cases}$$

với ràng buộc $\pi_1 + \pi_2 + \pi_3 = 1$. Sau khi giải hệ phương trình đơn giản này, chúng ta thu được phân phối dừng $\pi^{(\infty)}$:

$$\pi^{(\infty)} = \left[\frac{5}{23}, \frac{10}{23}, \frac{8}{23} \right]$$

B. Xích Markov thời gian liên tục

Tương tự như xích Markov thời gian rời rạc, quá trình này cũng yêu cầu về tính Markov. Có hai thành phần chính tạo nên một chuỗi Markov thời gian liên tục (CTMC): Đầu tiên cần có chuỗi Markov thời gian rời rạc còn được gọi là chuỗi nhảy (jump chain) cho xác suất chuyển P_{ij} ; Thứ hai, với mỗi trạng thái chúng ta có được tham số thời gian λ_i giúp kiểm soát lượng thời gian dành cho mỗi trạng thái thuộc $\mathcal{S} = \{1, 2, 3, \dots\}$ [9].

Định nghĩa 3: Ma trận sinh (generator matrix) còn được gọi là ma trận $\mathbf{G}$ hay ma trận $\mathbf{Q}$ giúp phân tích chuỗi Markov theo thời gian liên tục.

$$\mathbf{G} = \begin{bmatrix} g_{00} & g_{01} & g_{02} & \dots & g_{0j} & \dots \\ g_{10} & g_{11} & g_{12} & \dots & g_{1j} & \dots \\ g_{20} & g_{21} & g_{22} & \dots & g_{2j} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\ g_{i0} & g_{i1} & g_{i2} & \dots & g_{ij} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}$$

Ma trận $\mathbf{G}$ phải thỏa mãn những tính chất sau:

$$(i) \quad g_{ij} \geq 0 \quad \text{với mọi } i \neq j$$

$$(ii) \quad \sum_{j \in \mathcal{X}} g_{ij} = 0$$

$$(iii) \quad g_{ii} = -\sum_{j \neq i} g_{ij} = -\lambda_i$$

Xác định ma trận $\mathbf{P}(t)$ từ ma trận số mũ của ma trận $\mathbf{G}$ được định nghĩa như sau:

$$\mathbf{P}(t) = e^{\mathbf{G}t} = \sum_{n=0}^{\infty} \frac{\mathbf{G}^n t^n}{n!} \quad (14)$$

Vì $e^{\mathbf{G}0} = \mathbf{I}$, ta viết lại thành:

$$\mathbf{P}(t) = \mathbf{I} + \sum_{n=1}^{\infty} \frac{\mathbf{G}^n t^n}{n!} \Rightarrow \mathbf{P}(t) \approx \mathbf{I} + \mathbf{G}t + \frac{\mathbf{G}^2 t^2}{2!} \quad (15)$$

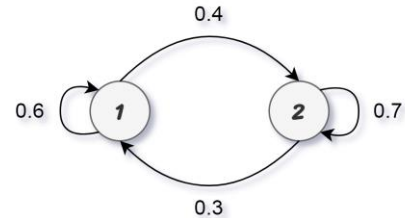
Phân bố xác suất tại thời điểm t thu được:

$$\pi^{(t)} = \pi^{(0)} \mathbf{P}(t) \quad (16)$$

III. MÔ PHỎNG SỐ

A. Thực nghiệm xích Markov thời gian rời rạc

Trong mục này chúng tôi sẽ xây dựng hai mô hình: một mô hình chuỗi Markov thông thường dựa trên ma trận chuyển tiếp với các giá trị xác suất được cho sẵn và mô hình ngăn SIS [10] có ma trận chuyển tiếp được tính dựa trên các trọng số aSI/N và bI dựa trên xích Markov thời gian rời rạc.



Hình 5: Chuỗi Markov với xác suất cho trước

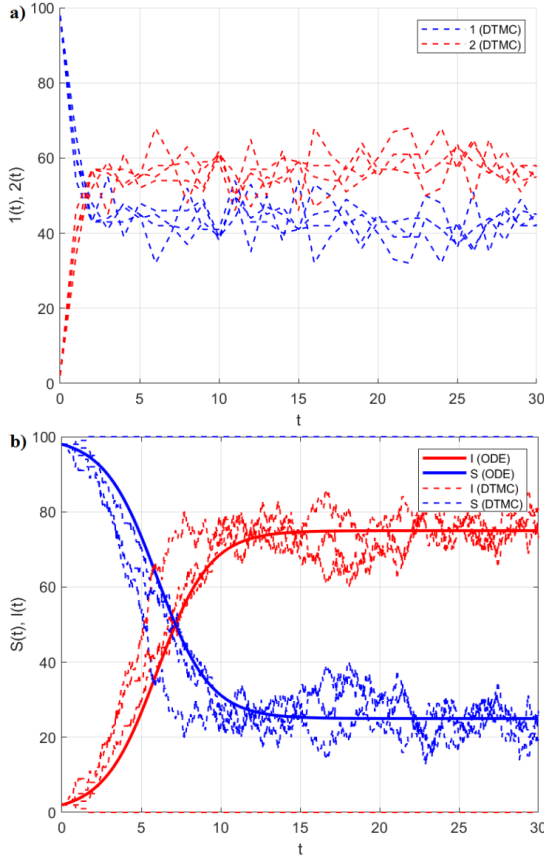
Cho chuỗi Markov như hình trên với khởi điểm $\pi^{(0)} = [98, 2]$ và ma trận xác suất chuyển cho mô hình thứ nhất:

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix} \end{matrix}$$

trong khi xác suất chuyển của mô hình thứ hai SIS phải thỏa mãn hệ sau:

$$p_{(S, I), (S+j, I+i)}(\delta t) = \begin{cases} \frac{aSI}{N} \delta t & (j, i) = (-1, 1) \\ bI \delta t & (j, i) = (1, -1) \\ 1 - \left(\frac{aSI}{N} + bI \right) \delta t & (j, i) = (0, 0) \\ 0 & (j, i) \text{ khác,} \end{cases} \quad (17)$$

với $S(0) = 98, I(0) = 2, a = 0.8$ và $b = 0.2$. Kết quả sau khi chạy thực nghiệm được mô tả trong hình 6:

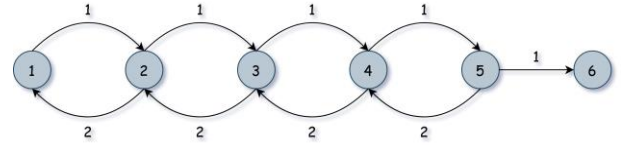


Hình 6: Mô phỏng số áp dụng phương pháp DTMC với $\text{sampling} = 4$: **a)** Mô hình chuỗi Markov thông thường, **b)** Mô hình SIS

Trong biểu đồ 6.a mô hình mô tả chuỗi Markov thông thường dần hội tụ dựa trên phân phối dừng $\pi^{(\infty)} = [0.4286, 0.5714]$. Về phần biểu đồ trong hình 6.b, ta có thể thấy được mô hình lấy mẫu có cả trường hợp mã độc tuyệt chủng ngay trong khoảng thời điểm đầu tiên trong khi mô hình tất định chỉ cho ra một kết quả duy nhất và không xét tới những yếu tố ngẫu nhiên bên ngoài.

B. Thực nghiệm xích Markov thời gian liên tục

Để có cái nhìn trực quan đầu tiên về Xích Markov thời gian liên tục, chúng tôi sẽ sử dụng chuỗi nhảy còn được gọi là quá trình nhảy (jump process) [16]. Giả sử rằng có một tập không gian trạng thái hữu hạn $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ với trạng thái thứ 6 còn được biết là trạng thái hấp thụ. Chúng tôi biểu diễn lược đồ các trạng thái như hình 7 dưới đây:

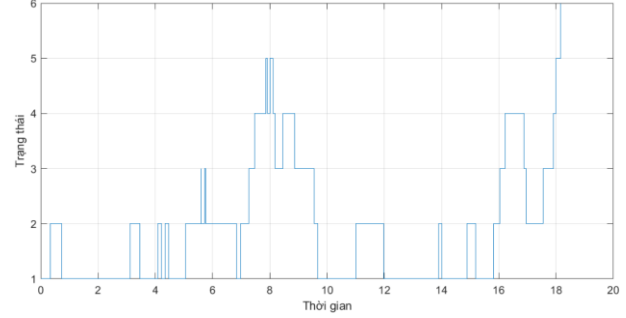


Hình 7: Chuỗi nhảy Markov

Ma trận $\mathbf{G}$ trong ví dụ này được biểu diễn như sau:

$$\mathbf{G} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 2 & -3 & 1 & 0 & 0 & 0 \\ 0 & 2 & -3 & 1 & 0 & 0 \\ 0 & 0 & 2 & -3 & 1 & 0 \\ 0 & 0 & 0 & 2 & -3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

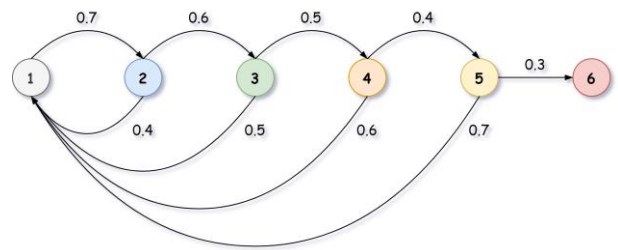
Dựa trên ma trận sinh cho trước, chúng tôi thực hiện mô phỏng số với kết quả thu được như hình sau:



Hình 8: Đồ thị biểu diễn quá trình mô phỏng quá trình nhảy với sinh mẫu là phân phối mũ có số ngẫu nhiên lấy từ phân phối đồng đều trên đoạn $[0, 1]$. Trạng thái dừng (6) đạt được tại sự kiện thứ 42

Từ quá trình nhảy chúng ta có thể mô phỏng lại quá trình mã độc cố gắng vượt các lớp bảo mật trong một khoảng thời gian dài, cụ thể là mã độc APT với cơ chế phòng thủ nhiều lớp [17] thường được áp dụng:

1. Lớp perimeter,
2. Lớp network,
3. Lớp host,
4. Lớp application,
5. Lớp data.

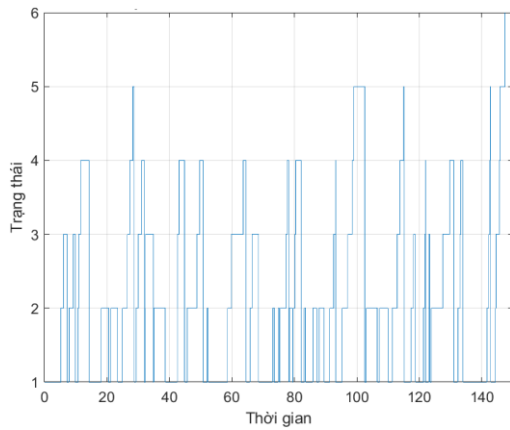


Hình 9: Xích Markov cho bài toán vượt các lớp bảo mật với giới hạn là 5 lớp

Ma trận sinh $\mathbf{G}$ được xây dựng dựa trên hình 9:

$$\mathbf{G} = \begin{bmatrix} -0.7 & 0.7 & 0 & 0 & 0 & 0 \\ 0.4 & -1 & 0.6 & 0 & 0 & 0 \\ 0.5 & 0 & -1 & 0.5 & 0 & 0 \\ 0.6 & 0 & 0 & -1 & 0.4 & 0 \\ 0.7 & 0 & 0 & 0 & -1 & 0.3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Kết quả thu được:



Hình 10: Đồ thị biểu diễn quá trình mô phỏng quá trình vượt qua các lớp bảo mật. Trạng thái dừng (6) đạt được tại sự kiện thứ 112

Bây giờ chúng tôi tiếp tục triển khai phương pháp xích Markov thời gian liên tục cho hai mô hình được dùng phổ biến cho bài toán dự đoán sự lan truyền của mã độc là: SI [13] và SIR [10,11], với xác suất chuyển trạng thái phải thỏa mãn các hệ sau.

- Đối với mô hình SI:

$$p_{(S, I), (S+j, I+i)}(\delta t) = \begin{cases} \frac{aSI}{N} \delta t + o(\delta t) & (j, i) = (-1, 1) \\ 1 - \frac{aSI}{N} \delta t + o(\delta t) & (j, i) = (0, 0) \\ o(\delta t) & (j, i) \text{ khác,} \end{cases} \quad (18)$$

- Đối với mô hình SIR:

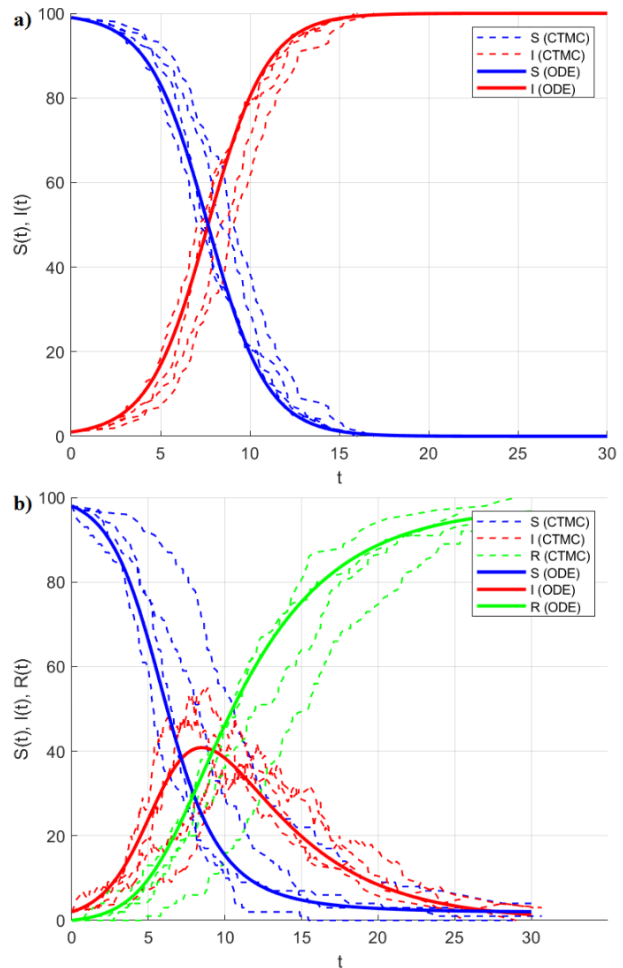
$$p_{(S, I), (S+j, I+i)}(\delta t) = \begin{cases} \frac{aSI}{N} \delta t + o(\delta t) & (j, i) = (-1, 1) \\ bI \delta t + o(\delta t) & (j, i) = (0, -1) \\ 1 - \left(\frac{aSI}{N} + bI \right) \delta t + o(\delta t) & (j, i) = (0, 0) \\ o(\delta t) & (j, i) \text{ khác,} \end{cases} \quad (19)$$

với mô hình SI: $S(0) = 99$, $I(0) = 1$ và $a = 0.6$; mô hình SIR: $S(0) = 98$, $I(0) = 2$, $R(0) = 2$, $a = 0.8$ và $b = 0.2$. Kết quả sau khi chạy thực nghiệm được mô tả trong hình 11.

So với các mô hình dựa trên thời gian rời rạc được biểu diễn ở hình 6, chúng ta thấy rằng đáng chú ý đồ thị trong hình 11 của những mô hình có thời gian liên tục có biên độ thay đổi nhỏ hơn sau mỗi một sự kiện diễn ra. Thông qua việc xây dựng các mô hình ngẫu nhiên từ hai phương pháp đã trình bày, ta có thể thấy được DTMC áp dụng tốt cho các bài toán phù hợp với bước thời gian như nhau, đòi hỏi ít dữ liệu ví dụ như dữ liệu thu thập được tính theo ngày hoặc theo tuần về số lượng máy bị nhiễm mã độc. Ngược lại, mô hình CTMC phù hợp với các bài toán có các sự kiện xảy ra liên tục ví dụ như tấn công DDOS, chính vì thế nó cũng đòi hỏi nhiều dữ liệu hơn so với DTMC.

IV. KẾT LUẬN

Trong bài báo này chúng tôi đã trình bày một cách ngắn gọn cơ sở lý thuyết cho hai quá trình ngẫu nhiên bao gồm xích Markov thời gian rời rạc và xích Markov thời gian liên tục. Tiếp đó, chúng tôi thực hiện mô phỏng số cho các mô hình ngẫu nhiên bao gồm SI, SIS và SIR dựa trên các phương pháp ngẫu nhiên đã trình bày ở mục 2 kết hợp với đầu ra



Hình 11: Mô phỏng số áp dụng phương pháp CTMC kết hợp sử dụng phương pháp Monte Carlo [18] với sampling = 4: a) Mô hình SI, b) Mô hình SIR

theo hướng tiếp cận dạng tắt định nhằm có cái nhìn cụ thể về sự khác biệt giữa hai loại mô hình này. Ngoài ra bài báo cũng mô phỏng thêm hai ví dụ về chuỗi Markov thông thường cho phương pháp DTMC và chuỗi nhảy cho phương pháp CTMC để có thể hiểu rõ hơn cách thức hoạt động của các quá trình ngẫu nhiên.

TÀI LIỆU THAM KHẢO

- [1] Minchin, E. A. (1909). The Prevention of Malaria. Journal of the Royal African Society, 8(31), 251–258. <http://www.jstor.org/stable/714977>.
- [2] Ross, R., & Hudson, H.P. (1917). An application of the theory of probabilities to the study of a priori pathometry.—Part I. Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences, 93, 204–230
- [3] Bacaër, N. (2011). Ross and malaria (1911). In: A Short History of Mathematical Population Dynamics. Springer, London. https://doi.org/10.1007/978-0-85729-115-8_12
- [4] Kermack, W.O., & McKendrick, A.G. (1927). "A Contribution to the Mathematical Theory of Epidemics". Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115(772), 700–721
- [5] Renshaw, E. 1993. Modelling Biological Population in Space and Time, Cambridge University Press
- [6] Fung, T., O'Dwyer, J.P., & Chisholm, R.A. (2019). Partitioning the effects of deterministic and stochastic processes on species extinction risk. Ecological Complexity

- [7] Shoemaker, L. G. L. L. Sullivan, I. Donohue, J. S. Cabral, R. J. Williams, M. M. Mayfield, J. M. Chase, C. Chu, W. Stanley Harpole, A. Huth, J. HilleRisLambers, A. R. M. James, N. J. B. Kraft, F. May, R. Muthukrishnan, S. Satterlee, F. Taubert, X. Wang, T. Wiegand, Q. Yang, and K. C. Abbott. 2020. Integrating the underlying structure of stochasticity into community ecology. *Ecology* 101(2):e02922. [10.1002/ecy.2922](https://doi.org/10.1002/ecy.2922)
- [8] O. Hernández-Lerma, J. B. Lasserre, Discrete-time Markov control processes: Basic optimality criteria, Springer Science & Business Media, ISSN 0172-4568, New York, NY, USA, 1996. <https://doi.org/10.1007/978-1-4612-0729-0>
- [9] Madhira, S., Deshmukh, S. (2023). Continuous Time Markov Chains. In: Introduction to Stochastic Processes Using R. Springer, Singapore. https://doi.org/10.1007/978-981-99-5601-2_6
- [10] Apostol, I. (2020). A Survey on Epidemiological Propagation Models of Botnets. *Journal of Military Technology*
- [11] A. Martín del Rey, R. Casado Vara, S. Rodríguez González, A computational propagation model for malware based on the SIR classic model, *Neurocomputing*, Volume 484, 2022, Pages 161-171, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2021.08.149>
- [12] Marc Artzrouni, Mathematical Demography, Editor(s): Kimberly Kempf-Leonard, Encyclopedia of Social Measurement, Elsevier, 2005, Pages 641-651, ISBN 9780123693983, <https://doi.org/10.1016/B0-12-369398-5/00360-1>
- [13] H. Ahn, J. Choi and Y. H. Kim, "A Mathematical Modeling of Stuxnet-Style Autonomous Vehicle Malware," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 673-683, Jan. 2023, doi: 10.1109/TITS.2022.3213771
- [14] José Roberto C. Piqueira, Manuel A.M. Cabrera, Cristiane M. Batistela, Malware propagation in clustered computer networks, *Physica A: Statistical Mechanics and its Applications*, Volume 573, 2021, 125958, ISSN 0378-4371, <https://doi.org/10.1016/j.physa.2021.125958>
- [15] Zhenhua Yu, Hongxia Gao, Dan Wang, Abeer Ali Alnuaim, Muhammad Firdausi, Almetwally M. Mostafa, SEI2RS malware propagation model considering two infection rates in cyber-physical systems, *Physica A: Statistical Mechanics and its Applications*, Volume 597, 2022, 127207, ISSN 0378-4371, <https://doi.org/10.1016/j.physa.2022.127207>
- [16] Bhattacharya, R., Waymire, E. (2023). Continuous Parameter Jump Markov Processes. In: Continuous Parameter Markov Processes and Stochastic Differential Equations. Graduate Texts in Mathematics, vol 299. Springer, Cham. https://doi.org/10.1007/978-3-031-33296-8_4
- [17] Abhishek Sisodia, (2021). Defense in Depth through Layered Security, Scientist – B, National Informatics Centre, A-Block, CGO Complex, Lodhi Road, New Delhi - 110003 India
- [18] Saidi Bidokhti, P. (Ed.). (2019). Theory, Application, and Implementation of Monte Carlo Method in Science and Technology. IntechOpen. doi: 10.5772/intechopen.78141

SOME TECHNIQUES FOR CONSTRUCTING STOCHASTIC MODELS

Abstract: In recent years, the surge in malware attacks has become a significant challenge, threatening many organizations, businesses, and individuals. To tackle this issue, network administrators have adopted mathematical models commonly used in epidemiology to track the trends

and intensity of malware attacks. These models can provide theoretical support for predictive decision-making and future malware control. Although numerous models have been proposed, researchers often favor deterministic models for their simplicity. However, these models are not always suitable, as unpredictable random events frequently occur, reducing the effectiveness of deterministic models. To address this limitation, some researchers have developed stochastic methods to enhance the realism of the models. Thus, in this paper, we aim to present several stochastic processes commonly applied in current stochastic models to improve cybersecurity applications.

Keywords: compartmental model, stochastic process, Markov chain.



Trần Hải Anh (1993). Tốt nghiệp đại học ngành Công nghệ thông tin tại Học viện Quản lý Giáo dục. Hiện đang là học viên cao học ngành Khoa học máy tính tại Học viện Công nghệ Bưu Chính Viễn Thông. Lĩnh vực nghiên cứu: giải tích số, mô hình hóa và mô phỏng, cấu trúc đồ thị, học máy thống kê, học sâu và hệ tư vấn.



Đỗ Xuân Chợt (1985). Nhận bằng thạc sĩ ngành Khoa học máy tính và cơ sở máy tính năm 2010, nhận bằng Tiến sĩ ngành Tự động hóa thiết kế tại Đại học Kỹ thuật điện quốc gia Saint Petersburg. Năm 2022, ông được bổ nhiệm làm phó giáo sư chuyên ngành an toàn thông tin và hiện đang là giảng viên Khoa An toàn thông tin Học viện Công nghệ Bưu chính Viễn thông và FPTU tại Việt Nam. Lĩnh vực quan tâm: bảo mật thông tin, mô hình hóa, hệ thống điều khiển, thuật toán hóa.