

Stochastic Modeling of Worm Propagation in Trusted Networks

Vivek Kumar Sehgal

Department of Electronics and Communication Engineering
Jaypee University of Information Technology,
Waknaghat, Solan – 173215 (HP) INDIA.

Abstract - There are two types of models useful in the study of worm propagation for a given number of terminals in a trusted network i.e. deterministic and stochastic model. The deterministic models, also known as compartmental models are in the form of epidemic models. These epidemic models consist of different states called compartments so these models are also known as compartmental models. A closed (no inflow and outflow) donor control based compartmental model can be converted in to the stochastic model which is more realistic and also enables us to compute the transition probability from one state to other state. In this paper authors have presented a method for stochastic modeling of worm or virus propagation in trusted networks.

Keywords: Epidemic SIR model, SEIR model, Compartmental matrix, Stochastic model, Transition probability matrix, Markov Chain.

1.0 Introduction

Over the past two decades computer worms have become worldwide problem. Worms can spread virtually moreover they are an ongoing, persistent, and worldwide problem on every popular micro computing platform. If we look at the past (two year) history we will find many cases like Morris worm in 1988 [1], Code Red appears on July 19, 2001 [2] which trigger the internet-scale worm attacks. In the same year Nimda infected thousands of computers [3]. In January 2003, the SQL Slammer worm was spread out and infected 90% valuable computers in very short period of time [4]. After that Cod Red II, Blaster, Sasser, and Witty have repeatedly attacked the internet. The history of these worms shows that how dangerous and how fast they spread to damage maximum computers in a network within fraction of time, before human can take effective precaution to prevent the damage of network.

The use of portable wireless devices like mobile phones, laptop, PDAs have been increased rapidly. Due to this mobile networks are becoming an important part of our everyday networking infrastructure which causes many cases of viruses in mobile network. For example, the Brador virus [5] infects Pocket PCs running Windows CE. The Cabir worm [6] infects cell phones running the Symbian operating system. The Mabir [7] and Symbos Comwar [8] worms use similar scanning techniques and also propagate via MMS messages.

This paper analyzes the spread of worm in a network with known number of hosts. The method of modeling is based on the epidemiology. Computer viruses and worms are similar to the biological viruses in their propagation behaviors. In epidemiology area, both stochastic models and deterministic models exist for modeling the spread of infectious diseases [9]. In this paper linear closed donor control based epidemic model is used as a linear compartmental (deterministic) model. From the compartmental model, we can derive the compartmental matrix which retains the properties of Metzler matrix. This compartmental matrix gives us the inter compartmental flow of closed epidemic model, which help us to drive the transition probability matrix and from that we can convert the compartmental matrix to the Markov Chain. The compartmental matrix used in this paper is mass conservative in the sense that mass balance is preserved inside the system. Pierre Philippe [10] shows some of the common models used in infectious disease modeling. In these models, some models are having feedback and some are not. The models with feedback can be converted in to stochastic models using Regular Markov Chains and the others can be converted into stochastic models using Absorbing Markov Chains. If the deterministic model is non linear compartmental model then it has to be linearized using Jacobian matrix about the equilibrium points. If the jacobian matrix is linear compartmental matrix then we can get the stochastic model.

1.1 Related work

Many researchers are working in the area of virus and worm modeling. To analyze the epidemics of these viruses and worms, Kephart, White, and Chess of IBM have performed a series of studies on viral infection based on epidemiology models [11]. Cliff C. Zou, Weibo Gong, Don Towsley, and Lixin Gao [12] present an Internet worm monitoring system and methodology to detect a worm at its early propagation stage by using Kalman filter estimation. Cliff C. Zou, Weibo Gong, and Don Towsley [13] presented the modeling and analysis under dynamic quarantine defense. Kurt Rohloff, Tamer Basar [14] worked on the detection of RCS worm epidemics. IBM's High Integrity Computing Laboratory demonstrates some interesting trends that have become apparent recently [15]. James W. Mickens, Brian D. Noble [16] has worked on modeling epidemic spreading in mobile environments

2.0 General Epidemic Model: Kermack–Mckendric Epidemic Model

In this section a basic SIR model is discussed where total number of hosts (N) is divided into three states (compartments) Therefore, in this model each host stays in any one of the three states at any instant of time. The three states are: susceptible (S), infectious (I), or removed (R). A host either makes the state transition “susceptible \rightarrow infectious \rightarrow removed” or remains in the susceptible compartment as shown in the Figure 1.

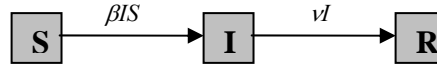


Figure 1: Simple SIR epidemic model

The model is a nonlinear epidemic model where βIS component is the source of non linearity. Therefore the compartmental model is also a nonlinear. This model can be described by the following set of equations:

$$dS/dt = -\beta IS \quad (1)$$

$$dI/dt = \beta IS - \nu I \quad (2)$$

$$dR/dt = \nu I \quad (3)$$

$$N = S + I + R \quad (4)$$

Where β is the *transmission* rate and ν is the *removal* rate of infectious hosts [13]. The *epidemic threshold* of this model is:

$$\rho \equiv \nu/\beta \quad (5)$$

A major outbreak occurs if and only if the initial number of susceptible hosts $S(0) > \rho$. From equations (1) to (4) we can derive $dI/dt < 0, \forall t > 0$ if and only if $S(0) < \rho$. This model can be used as deterministic model which is nonlinear in nature.

Therefore it has to linearize using Jacobian matrix J about equilibrium points (S_e, I_e)

$$J = \begin{bmatrix} -\beta I_e & -\beta S_e & 0 \\ \beta I_e & \beta S_e - \nu & 0 \\ 0 & \nu & 0 \end{bmatrix} \quad (6)$$

This Jacobian matrix (6) is not a compartmental matrix, so this model can not be converted in to stochastic model. We can have another epidemic model called SEIR model, which is linear and can be converted in to the stochastic model. Figure 2 shows the dynamic behavior of basic SIR model for $N = 105$, $S(0) = 100$, $I(0) = 5$, and $R(0) = 0$. The transmission rate $\beta = 0.001$ and removal rate $\nu = 0.05$.

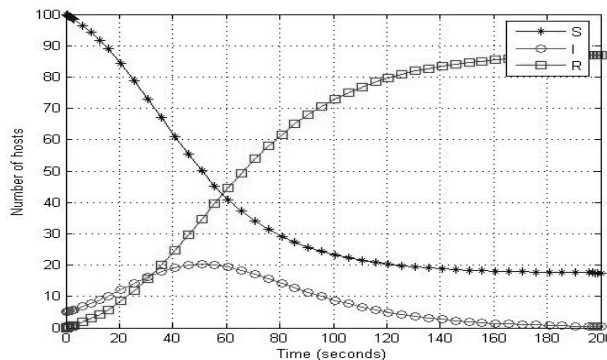


Figure 2: Dynamic behavior of basic SIR model

3.0 Epidemic Model for Stochastic Modeling

In this section a linear epidemic model SEIR with four compartments is discussed. These compartments are: susceptible (S), exposed (E), infectious (I), and removed (R). This model can be used for stochastic modeling with following certain assumptions:

- 1) Total number of hosts is constant.
- 2) Compartment E is merged with compartment I.
- 3) The hosts in compartment E are directly recovered.
- 4) The model is donor control based model.
- 5) The model is mass conservative.

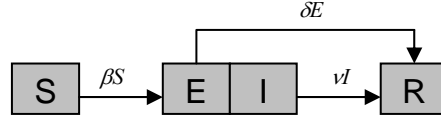


Figure 3: Linear SEIR epidemic model

A host either makes the state transition “susceptible \rightarrow exposed \rightarrow removed” or “susceptible \rightarrow infectious \rightarrow removed” as shown in the Figure 3. The hosts in compartment (E) are infected or exposed but not yet infectious at time t . The behavior of SEIR model in Figure 3 can be described by the following set of differential equations:

$$\frac{dS}{dt} = -\beta S \quad (7)$$

$$\frac{dE}{dt} = -\delta E + (1 - \alpha)\beta S \quad (8)$$

$$\frac{dI}{dt} = -\nu I + \alpha\beta S \quad (9)$$

$$\frac{dR}{dt} = \delta E + \nu I \quad (10)$$

$$N = S + E + I + R \quad (11)$$

Where β is the *transmission* rate and ν, δ are the *removal* rates of infected and infectious hosts, α is the separation constant. This model can be used as deterministic model. Figure 4 shows the dynamic behavior of basic SEIR model for $N = 15000, S(0) = 14000, E(0) = 50, I(0) = 950$ and $R(0) = 0$. The transmission rate $\beta = 0.01$ and removal rates $\nu = 0.01, \delta = 0.01$. The separation constant α is 0.001. Since the equations describing the behavior of SEIR model are linear differential equations, the four compartments appearing in these equations can be treated as physical state space variables.

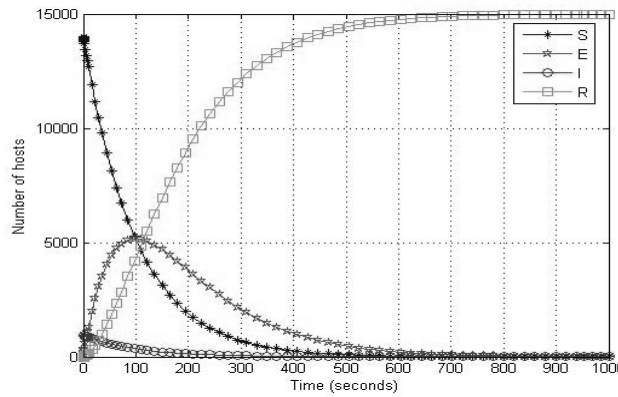


Figure 4: Dynamic behavior of basic SEIR model

4.0 Compartmental Modeling of SEIR Model

The state space model of SEIR epidemic model can be described by the following equations:

$$\dot{X}(t) = AX(t) \quad (12)$$

$$\begin{bmatrix} \dot{S} \\ \dot{E} \\ \dot{I} \\ \dot{R} \end{bmatrix} = \begin{bmatrix} -\beta & 0 & 0 & 0 \\ (1-\alpha)\beta & -\delta & 0 & 0 \\ \alpha\beta & 0 & -\nu & 0 \\ 0 & \delta & \nu & 0 \end{bmatrix} \begin{bmatrix} S \\ E \\ I \\ R \end{bmatrix} \quad (13)$$

$$A = \begin{bmatrix} -\beta & 0 & 0 & 0 \\ (1-\alpha)\beta & -\delta & 0 & 0 \\ \alpha\beta & 0 & -\nu & 0 \\ 0 & \delta & \nu & 0 \end{bmatrix} \quad (14)$$

Where A is called compartmental matrix. The solution of this homogeneous state equation is:

$$X(t) = e^{At} X(0) \quad (15)$$

$$X(t) = \begin{bmatrix} S(t) \\ E(t) \\ I(t) \\ R(t) \end{bmatrix} \quad (16)$$

$$e^{At} = L^{-1} \left[(sI - A)^{-1} \right] \quad (17)$$

$$\text{or } e^{At} = I + At + \frac{1}{2!} A^2 t^2 + \dots + \frac{1}{i!} A^i t^i \quad (18)$$

Where e^{At} called state transition matrix of SEIR is deterministic model and $X(0)$ is the column matrix which shows the initial conditions of SEIR epidemic model

4.1 Properties of Compartmental Matrix

Certain important properties of compartmental matrix A are given below:

- 1) The diagonal elements of compartmental matrix are zero or negative elements.
- 2) The non-diagonal elements of compartmental matrix are zero or positive.
- 3) The first eigenvalue of compartmental matrix is zero.
- 4) The sum of elements in each column of compartmental matrix is equal to zero.
- 5) Compartmental matrix is Metzler matrix.

5.0 Stochastic Modeling of SEIR Model

Stochastic modeling is very useful in worm epidemic model. Stochastic model help us to calculate the transition probability from one state to other state and expected time from one state to other state. The transition probability matrix can be derived from compartmental matrix using following relation [17].

$$P = (I + hA)^T \quad (19)$$

Where P is the transition probability matrix and h is the time between events or trials.

Proof:

The probability $p_i(n)$ that the random variable is in state i at any time n may be found from the level of numbers or quantity of random variables $x_i(n)$ in that state (now called compartment) at time n . Indeed $p_i(n) = x_i(n) / \sum_{j=1}^k x_j(n)$, where k is the number of states. The levels at time $n+1$ are given in terms of those at time n by the same equation,

$$X_{n+1}^T = X_n^T P, \quad n = 0, 1, 2, \dots \quad (20)$$

as the probabilities. Here, X_n is a column vector of material levels. Then, we have

$$X_{n+1}^T \begin{bmatrix} 1 \\ 1 \\ \cdot \\ 1 \end{bmatrix} = \begin{bmatrix} x_{n+1,1} \cdot x_{n+1,2} \cdot \dots \cdot x_{n+1,k} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \cdot \\ 1 \end{bmatrix} = X_n^T P \begin{bmatrix} 1 \\ 1 \\ \cdot \\ 1 \end{bmatrix} = X_n^T \begin{bmatrix} 1 \\ 1 \\ \cdot \\ 1 \end{bmatrix} \quad (21)$$

Since $[1, 1, \dots, 1]^T$ is always a right eigenvector corresponding to the steady state eigenvalue of 1 of P . If we started with a

quantity $q = \sum_{j=1}^n x_j(0)$ of materials in the system, then the total quantity in the system remains at q for all time by (21).

Thus, we have $p_i(0) = \frac{x_i(0)}{q}$.

Thus, equation (20) is one form of equation of a compartmental system, but a more common format is as a difference equation

$$X_{n+1}^T - X_n^T = X_n^T (P - I)$$

or by taking transpose of the matrices

$$\Delta X_n = (P^T - I) X_n \quad (22)$$

If the time step, i.e., the time between trials, is h rather than 1, then $X_n = X(nh)$ and the left side of equation (22) is replaced by the difference quotient

$$\frac{X(nh+h) - X(nh)}{h} = \frac{1}{h} (P^T - I) X(nh) := AX(nh)$$

We now let $t = nh$ to get

$$\frac{X(t+h) - X(t)}{h} = AX(t)$$

This left side is approximately the derivative, so we have

$$X' = AX$$

This is the differential equation for the compartmental matrix. Hence

$$A = \frac{1}{h} (P^T - I)$$

Where P is the transition probability matrix and h is the time between events or trials. Hence $P = (I + hA)^T$.

$$P = \begin{bmatrix} 1-h\beta & 0 & 0 & 0 \\ h(1-\alpha)\beta & 1-h\delta & 0 & 0 \\ h\alpha\beta & 0 & 1-h\nu & 0 \\ 0 & h\delta & h\nu & 1 \end{bmatrix}^T = \begin{bmatrix} p_{ss} & p_{se} & p_{si} & p_{sr} \\ p_{es} & p_{ee} & p_{ei} & p_{er} \\ p_{is} & p_{ie} & p_{ii} & p_{ir} \\ p_{rs} & p_{re} & p_{ri} & p_{rr} \end{bmatrix} = \begin{bmatrix} 1-h\beta & h(1-\alpha)\beta & h\alpha\beta & 0 \\ 0 & 1-h\delta & 0 & h\delta \\ 0 & 0 & 1-h\nu & h\nu \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (23)$$

From (23) we can see that the sum of all elements in each row of transition probability matrix P is equal to 1. Hence

$$\sum_{j=0}^{j=3} P_{ij} = 1 \quad \text{Where } i, j=0 \dots 4 \quad (24)$$

5.1 Properties of Transition Probability Matrix

Certain important properties of transition probability matrix P are given below:

- 1) The first eigenvalue of transition probability matrix is one.
- 2) The sum of all elements in each row of transition probability matrix is equal to 1.
- 3) This matrix is also known as Markov Matrix.

5.2 Markov Chain from Transition Probability Matrix

Figure 5 shows the stochastic diagram of transition probability matrix P

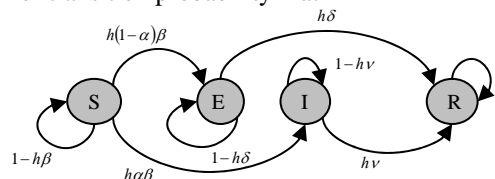


Figure 5: Stochastic diagram (Absorbing Markov Chain) of SEIR model

In an Absorbing Markov Chain with states ordered such that the transition probability matrix P has the form

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix} \quad (25)$$

the following hold:

- 1) $Q^t \rightarrow 0$ as $t \rightarrow \infty$
- 2) $R_\infty = (1-Q)^{-1}R$
- 3) The expected number of times a chain is in the nonabsorbing state k_j given that it started in k_i is given by the corresponding element of $(1-Q)^{-1}$. The matrix $(1-Q)^{-1}$ is often referred to as *Markov chain's fundamental matrix*

For each nonabsorbing state, there is an absorbing state with a path of minimum length. Let r be the maximum length of all such paths. Therefore, in r steps, there is a positive probability p of entering one of the absorbing states regardless of where you started. The probability of not reaching an absorbing state in r steps is $(1-p)$. After the next r steps, it is $(1-p)^2$ and after kr steps, $(1-p)^k$. Since this approaches 0 as $k \rightarrow \infty$, the probability of being in any nonabsorbing state approaches 0 as $t \rightarrow \infty$. But the elements of Q^t are just these probabilities.

6.0 Stochastic Analysis of SEIR Model

From equations (23) and (25), we get

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1-h\beta & h(1-\alpha)\beta & h\alpha\beta \\ 0 & 1-h\delta & 0 \\ 0 & 0 & 1-h\nu \end{bmatrix} & \begin{bmatrix} 0 \\ h\delta \\ h\nu \end{bmatrix} \\ \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 \end{bmatrix} \end{bmatrix} \quad (26)$$

The last state of this Markov chain $p_{rr} = 1$, is absorbing state. For $\beta = \delta = \nu = 0.01$, $\alpha = 0.001$ and $h = 0.1$

$$P = \begin{bmatrix} 0.999 & 0.000999 & 0.000001 & 0 \\ 0 & 0.999 & 0 & 0.001 \\ 0 & 0 & 0.999 & 0.001 \\ 0 & 0 & 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} 0.999 & 0.000999 & 0.000001 \\ 0 & 0.999 & 0 \\ 0 & 0 & 0.999 \end{bmatrix}, R = \begin{bmatrix} 0 \\ 0.001 \\ 0.001 \end{bmatrix}, I = 1$$

For transient response

$$(1-Q)^{-1} = \begin{bmatrix} 1000 & 999 & 1 \\ 0 & 1000 & 0 \\ 0 & 0 & 1000 \end{bmatrix}$$

From $(1-Q)^{-1}$ matrix we can calculate:

- 1) For susceptible host
Expected time as susceptible host = $(1-Q)^{-1}_{11} = 1000$ Sec. Expected time as exposed host = $(1-Q)^{-1}_{12} = 999$ Sec. Expected time as infectious host = $(1-Q)^{-1}_{13} = 1$ Sec.
Total expected time that a susceptible host takes to become infectious = $1000+999+1 = 2000$ Sec.
- 2) For exposed host
Expected time as exposed host = $(1-Q)^{-1}_{22} = 1000$ Sec.
- 3) For infectious host
Expected time as infectious host = $(1-Q)^{-1}_{33} = 1000$ Sec.

The expected time for exposed host as infectious host = 0, because there is no transition from E state to I state as shown in Figure 3

For steady state response

$$R_\infty = (1-Q)^{-1}R = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

From $R_\infty = (1-Q)^{-1}R$ matrix we can calculate:

- 1) Probability of susceptible host to recover = $(1-Q)^{-1}R_{11} = 1$
- 2) Probability of exposed host to recover = $(1-Q)^{-1}R_{21} = 1$
- 3) Probability of infectious host to recover = $(1-Q)^{-1}R_{31} = 1$

For the steady state complete transition matrix $P_\infty = \begin{bmatrix} Q_\infty & R_\infty \\ 0 & I \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

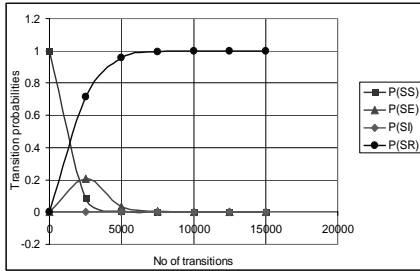


Figure 6: Transition Prob. for state S

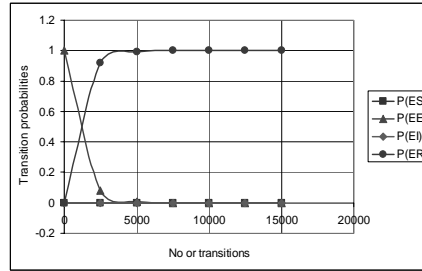


Figure 7: Transition Prob. for state E

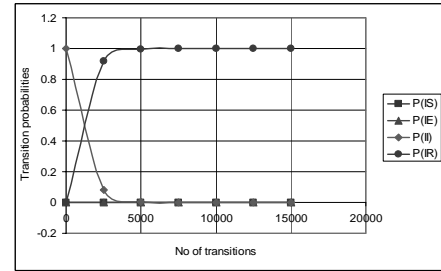


Figure 8: Transition Prob. for state I

7.0 Conclusion

The spread of worm or virus in a network is a random process and this paper help us to find out the various transition probabilities and expected time of state transition so we can inject the antivirus in time. Further extension of this work is to design the stochastic model for a epidemic model with feedback.

This work has good scope in future. If we take the epidemic model with the inflow and outflow rates then we can check the controllability and absorability of epidemic model, which help us to design the state observer and state space controller to control the spread of worm. This state space controller is also called compartmental controller.

8.0 References

- [1] D. Seeley, "A tour of the worm," in Proc. Winter USENIX Conf., Jan.1989, pp. 287–304.
- [2] D. Moore, C. Shannon, and J. Brown, "Code-Red: A case study on the spread and victims of an Internet worm," in Proc. 2nd ACM SIGCOMM Workshop on Internet Measurement, Nov. 2002, pp. 273–284.
- [3] CAIDA. Dynamic Graphs of the Nimda worm. <http://www.caida.org/dynamic/analysis/security/nimda/>
- [4] D. Moore, V. Paxson, S. Savage, C. Shannon, S.Stanford, and N. Weaver. Inside the Slammer Worm. IEEE Security and Privacy, 1(4):33-39, July 2003.
- [5] R. Wong and I. Yap. Security Information: Virus Encyclopedia: WINCE BRADOR.A: Technical Details, 2004. Trend Micro Incorporated.
- [6] P. Ferrie, P. Szor, R. Stanev, and R. Mouritzen. Security Response: SymbOS.Cabir, 2004. Symantec Corporation.
- [7] E. Chien. Security Response: SymbOS.Mabir, 2005. Symantec Corporation.
- [8] M. Lactaotao. Security Information: Virus Encyclopedia: SYMBOS COMWAR.A: Technical Details, 2005. Trend Micro Incorporated.
- [9] D.J. Daley and J. Gani. Epidemic Modelling: An Introduction. Cambridge University Press, 1999.
- [10] <http://www.ispub.com/istia/index.php?xmlFilePath=journals/ijid/vol1...>
- [11] J. O. Kephart and S. R. White. Directed-graph Epidemiological Models of Computer Viruses. In Proceedings of the IEEE Symposium on Security and Privacy, 1991.
- [12] Cliff C. Zou, Weibo Gong, Don Towsley, and Lixin Gao "The Monitoring and Early Detection of Internet Worms," IEEE/ACM Trans. Networking.
- [13] Cliff C. Zou, Weibo Gong, and Don Towsley. "Worm Propagation Modeling and Analysis under Dynamic Quarantine Defense," ACM CCS Workshop on Rapid Malcode (WORM'03), Oct. 27, Washington DC, USA, 2003.
- [14] Kurt Rohloff, Tamer Basar "The detection of RCS worm epidemics" Proc. 2005 ACM workshop on Rapid malcode WORM '05
- [15] <http://www.research.ibm.com/antivirus/SciPapers/White/VB95/vb95.distrib.html>
- [16] James W. Mickens, Brian D. Noble "Modeling epidemic spreading in mobile environments" Proc. 4th ACM workshop on Wireless security WiSe '05
- [17] Gilbert G. Walter and Martha Contreras "Compartmental Modeling with Networks"
- [18] G.S. Ladde, "Cellular Systems - II. Stability of Compartmental Systems", Mathematical Biosciences, 30, pp. 1-21, 1976.
- [19] J.A. Jacquez, C.P. Simon, "Qualitative Theory of Compartmental Systems with Lags", Mathematical Biosciences, vol. 180, pp. 329 - 362, 2002.
- [20] I.W. Sandberg, "On the Mathematical Foundations of Compartmental Analysis in Biology, Medicine, and Ecology", IEEE Trans. Circuits and Systems, CAS25(5), pp. 273-279, 1978.