Queueing Network Realization of an Epidemiological Model for Efficient Evaluation of Computer Transmitted Infections

N. Eva Wu, Joshua Montague, Drake W. Van Ornam, Morteza Sarailoo, and John S. Bay

Department of Electrical and Computer Engineering, Binghamton University, SUNY, Binghamton, NY 13902, USA
(Tel: 607-777-4375; e-mail:evawu,jmontag2,dvanorn1,msarail1,bayj@binghamton.edu).

Abstract: This paper reexamines an epidemiological model with 4 population groups (vigilant/non-vigilant susceptible/infectious) built to study the effect of user vigilance on computer transmitted infections (CTIs) in computer networks. The model serves as an example through which a model conversion process is delineated, which aims at enhancing computational efficiency in the evaluation of the global prevalence of CTIs. More specifically, the conventional node-centric networked Markov chain (NCMC) is remodeled as a population-centric Markov chain (PCMC) to reduce the state-space size from an exponential to a polynomial function of the number of computing nodes N in a strongly connected network, where external attack and internal spread processes are aggregated. The PCMC is then realized as a closed queueing network of 4 M/M/N/N queueing nodes, corresponding to the 4 population groups. The results of evaluating the evolution of mean populations for the 4-population network of up to 150,000 computing nodes show that the queueing network realization slows the growth of computational complexity from exponential to linear with respect to the network size without resorting to mean field approximations. The paper briefly discusses on how the queueing network framework can accommodate node-centric Markov chains (NCMCs) of arbitrary directed networks of heterogeneous nodes, and its potential to significantly reduce the complexity in the evaluation of mean population dynamics for the more general class of large networks.

Keywords: computer transmitted infections, epidemiological models, Markov models, computational methods, closed queueing networks, performance evaluation, and probabilistic risk assessment.

1. INTRODUCTION

Epidemiological models introduced nearly a century ago by Kermack and McKendrick [6] have been used to study equilibria as well as transient behaviours of computer transmitted infections (CTI) in computer networks, and more recently to serve as the bases of control designs for CTI mitigation [9]. This paper seeks to develop a stochastic modelling and computation framework for more efficient evaluation of mean infectious population, and ultimately, cost-constrained control of CTIs under the new framework.

The proposed queueing network formalization here builds on a sample 4-population model of Kelley and Camp (K-C) [5], in which $(S_r, S_a, I_r, I_a)(t)$ represents the evolution of non-vigilant susceptible, vigilant susceptible, non-vigilant infectious, and vigilant infectious populations, respectively, in a network of N computing nodes. We interpret subscript r as implying *r*isk-seeking (non-vigilant), and a as risk-*a*version (vigilant) users. The K-C paper sought to determine from data the parameters of a non-vigilant/vigilant SIS (susceptible-infectious-susceptible) model (N/VSIS hereafter), and study the sensitivity of the global prevalence of CTIs, quantified by the steady-state fraction of the total infectious population, with respect to the model parameters, especially those capturing the effects of patching and upgrading, as well as targeted social pressure [15]. Similar studies to model and understand the effects of user awareness/alert can be found in [10], [13], and some references therein.

The N/VSIS network [5] is selected here to represent a class of networks, whose epidemiological models have been linked through mean-field approximations to higher dimensional Markov chains. A mean-field approximation [14] replaces the random population-dependent state transition rates by an expectation with respect to node-specific marginal distributions. The class contains variations and extensions of the Kephart and White homogeneous 2-population SIS mean-field model [7], [14]. Examples in this class include the 3-population SIRS (R means recovered) network [12], for which the relationship between a 3^N-state networked Markov chain and its $3N$-state mean-field approximation was explored, and the 4-population G-SEIV (G means generalized and E means exposed) network [9], whose mean-field approximation provided the basis for optimal allocation of resources to speed up the eradication of the infectious populations. These models typically represent networks with a small number, m, of population groups, typically a low single digit number, and a large number, N, of computing nodes. The curse of dimensionality in a large network limits the utility of Markov chains without mean-field approximations, whose validity and accuracy, however, are not universally held. In the case of the N/VSIS network, for example, a mean-field-like model was a least-squares fit of a parametric system of differential equations to a single sample path of attack data [5].

This paper challenges the conventional modelling practice for CTIs, and aims to enable a more efficient numerical evaluation of $E(S_r, S_a, I_r, I_a)(t)$, the mean population dynamics, without resorting to mean-field approximations, and pave the way for

their optimal control [1]. The 4-population N/VSIS network serves as a vehicle to explain the significant computational advantage of the proposed queueing network realization of a proposed population-centric Markov chain (PCMC) in relation to the existing node-centric networked Markov chains (NCMC). Terms NCMC and PCMC are coined in this paper and are contrasted in multiple occasions hereafter. More specifically, the dimension of the PCMC state-space is $(N + 3)(N + 2)(N + 1)/6$, a polynomial, reduced from 4^N, an exponential function in an NCMC, for an N-node network. This paper further implements a PCMC as a closed queueing network [3] with 4 M/M/N/N queueing nodes (Poisson inter-arrival time/exponential service time/N-server per queueing node/N-entity network). The paper establishes the validity of the queueing network formalization of the PCMC, and verifies their equivalence numerically. The results of evaluating the mean population evolution for networks of 15 to 150,000 nodes show that the queueing network realization of a PCMC leads to a linear growth of computational time with respect to the growth of network size N. Finally, the extension of the queueing network modelling framework is discussed briefly to accommodate networks of heterogeneous nodes and node interactions.

The paper is organized as follows. Section 2 interprets the N/VSIS model as a mean-field approximation of a more detailed NCMC, states the assumptions used in the modified N/VSIS and constructs its PCMC. Section 3 formalizes a queueing network realization of the PCMC, demonstrates its computational advantage via evaluating mean population evolution for networks with up to 150K nodes, and describes several extensions. Section 4 concludes the paper and briefly discusses the planned future research.

2. N/VSIS NETWORK AS NCMC AND PCMC

In this section, the Kelley and Camp N/VSIS model [5] is re-examined. Its original form is reinterpreted as a mean-field-like approximation of a more rigorously built node-centric Markov chain (NCMC), whose construction is briefly discussed. We then propose to reconstruct the Markov chain with a population-centric approach (PCMC) in which the infection process of a node is triggered by an aggregated event of internal spread and external attack as an independent Poisson arrival process. The goal is to use the modified 4-population N/VSIS as a representative network to elucidate the distinction and the relation between the existing NCMC and the proposed PCMC modelling frameworks that can be generalized to any m-population N-node networks.

2.1 N/VSIS and its mean-field-like approximation

The flow of populations $(S_r, S_a, I_r, I_a)(t)$ was expressed in [5] by four dependent differential equations. Equation (1) below, where $S_r(t) + S_a(t) + I_r(t) + I_a(t) = N$, slightly modifies the K-C N/VSIS model to uphold the flow balance principle [3].

$$\frac{dS_r}{dt} = -(\eta + \beta_r)\frac{I_r + I_a}{N}S_r + \delta S_a + \mu_r I_r$$
$$\frac{dS_a}{dt} = \eta\frac{I_r + I_a}{N}S_r - (\delta + \beta_a\frac{I_r + I_a}{N})S_a + (\mu + \gamma\frac{S_a}{N})I_r + (\mu_a + \gamma_a\frac{S_a}{N})I_a$$
$$\frac{dI_r}{dt} = \beta_r\frac{I_r + I_a}{N}S_r - (\mu_r + \mu + \gamma\frac{S_a}{N})I_r$$
$$\frac{dI_a}{dt} = \beta_a\frac{I_r + I_a}{N}S_a - (\mu_a + \gamma_a\frac{S_a}{N})I_a$$

(1)

The original notations used in [5] are largely retained in (1). They are defined in Table 1 as transition rates of a single network node between two population groups for use later in the stochastic versions of the model. The rate values in the table are set to reflect the N/VSIS utility with $\beta_r > \beta_a > 0$ implying that a vigilant node is less prone to attacks, $0 < \mu_r < \mu_a$ implying that a vigilant node is quicker to recover, and η, $\mu > 0$, representing the effort to promote vigilance. Note that η, β_r, β_a, γ_a, and γ in (1) have population variables attached to them that cause the model to be nonlinear with bifurcation and epidemic threshold [9],[14] indicating whether the network can be infection-free in the long run.

Table 1 Transition rates in Equation (1) (1/unit time)

Symbol	Transition rate of a single node between two populations
η	Upgrade: non-vigilant susceptible to vigilant susceptible
δ	Downgrade: vigilant susceptible to non-vigilant susceptible
β_r	Infect: non-vigilant susceptible to non-vigilant infectious
β_a	Infect: vigilant susceptible to vigilant infectious
μ_r	Recover&upgrade: non-vigilant infectious to vigilant susceptible
μ_a	Recover: vigilant infectious to vigilant susceptible
μ	Recover: non-vigilant infectious to vigilant susceptible
γ	Social pressure: non-vigilant infectious to vigilant susceptible
γ_a	Social pressure: vigilant infectious to vigilant susceptible

The underlying N-node network of the N/VSIS model (2) is understood as being strongly connected and homogeneous. The homogeneity assumption recognizes random scanning as a common online attack mechanism rendering well-mixed populations in certain sociotechnical networks, differing from most biological epidemics [11]. The construction of a 4^N-state node-centric networked Markov chain (NCMC) for the 4-population N-node N/VSIS network can follow the procedure described in [9] for a 4-population N-node heterogeneous G-SEIV network. With some abuse of notations, we now interpret (S_r, S_a, I_r, I_a) as state names in the stochastic setting for an NCMC associated with a generic node. Each node in this NCMC can be in any one of the four states at any given time with a certain probability. A mean-field approximation of the NCMC can be sought in the form of a set of 4N-dependent ordinary differential equations on marginal probability distributions $(p_{S_r}^i, p_{S_a}^i, p_{I_r}^i, p_{I_a}^i)(t)$ for node i, where $i = 1, \cdots, N$. Homogeneity further reduces a 4N-variable system to the 4-variable mean-field-like approximation below, where $p_{S_r}(t) + p_{S_a}(t) + p_{I_r}(t) + p_{I_a}(t) = 1$.

$$\frac{dp_{S_r}}{dt} = -(\eta + \beta_r)(p_{I_r} + p_{I_a})p_{S_r} + \delta p_{S_a} + \mu_r p_{I_r}$$
$$\frac{dp_{S_a}}{dt} = \eta(p_{I_r} + p_{I_a})p_{S_r} - \delta p_{S_a} - \beta_a(p_{I_r} + p_{I_a})p_{S_a}$$
$$+ \mu p_{I_r} + \gamma p_{S_a}p_{I_r} + \mu_a p_{I_a} + \gamma_a p_{S_a}p_{I_a}$$

(2)

$$\frac{dp_{I_r}}{dt} = \beta_r(p_{I_r} + p_{I_a})p_{S_r} - (\mu_r + \mu)p_{I_r} - \gamma p_{S_a}p_{I_r}$$
$$\frac{dp_{I_a}}{dt} = \beta_a(p_{I_r} + p_{I_a})p_{S_a} - \mu_a p_{I_a} - \gamma_a p_{S_a}p_{I_a}$$

System (2) governs the evolution of population probabilities of the underlying stochastic process, whereas system (1) governs the evolution of the population flow, and is therefore deterministic. Note that (2) so far is merely a conjecture using a similar argument to that provided in [14], where it only concluded the validity of the early Kephart-White epidemic model [7] as a mean-field approximation of a 2^N-state NCMC for a 2-population SIS network. The earlier reference to N/VSIS model (1) as a mean-field-like approximation can be explained by a comparison of (2) and rescaled (1) in which all population variables are normalized by total population N.

2.2 Population-centric Markov chains for N/VSIS

We now reconstruct a Markov chain for the strongly connected and homogeneous N/VSIS network, in which independent Poisson event clocks define all interevent time distributions [3] for transitions between population groups of any single nodes. Two major departures distinguish our proposed stochastic model from the NCMCs in [14], [12], and [9].

(A1) The state space is modified from node-centric

$$x = (x_1, \cdots x_N) \in S_{NC}, \ x_i \in \{S_r, S_a, I_r, I_a\}, \ i = 1, \cdots, N \quad (3)$$

to population-centric

$$x = (S_r, S_a, I_r, I_a) \in S_{PC}, \ S_r, S_a, I_r, I_a \in \{0, 1, \cdots, N\},$$
$$S_r + S_a + I_r + I_a = N. \quad (4)$$

(A2) Infection process of a susceptible node is triggered by an aggregated Poisson event arrival representing both internal spread and external attack. Thus, infection rates β_r, β_a, and upgrade rate η are assumed to be unaffected by the infectious populations. Moreover, the social response in [5] are ignored by setting $\gamma = \gamma_a = 0$ to focus on our goal to develop PCMCs. Point of departure (A1) dictates the size of the state space defined in (4) for a PCMC as a function of total population N

$$|S_{PC}| = (N+3)!/N!/3! \sim O(N^3), \quad (5)$$

which is reduced from $|S_{NC}| = 4^N$ for an NCMC whose state space is defined in (3). Point of departure (A2) leads to irreducible Markov chains [3] for both the NCMC and the PCMC. As a result, there no longer exists an infection-free steady-state distribution.

Remark 1. Lack of uniform fit of parametric N/VSIS model (1) to a sample path of attack data [5] suggests that the assumed N/VSIS structure may be too restrictive to properly capture the population dynamics $(S_r, S_a, I_r, I_a)(t)$ in modern networks. A better fit can be expected in the determination of the transition rates in Table 2 that define individual interevent time distributions using, say, maximum likelihood estimation and goodness of fit tests [4] with labelled data for each triggering event.

To gain a concrete sense of the relation between an NCMC and a PCMC, we list in Table 2 the states in their respective state spaces for a 2-node N/VSIS network. Symbols S_r, S_a, I_r, and I_a are again somewhat abused as they are defined differently for the PCMC in the 1st column and for the NCMC in the 2nd column of the table. There are $5!/(2! \times 3!) = 10$ states in the PCMC, whereas there are $m^N = 4^2 = 16$ states in the NCMC. A partial rate transition diagram of a population-centric Markov chain (PCMC) built for the modified N/VSIS network is shown in Fig.1, where state transition rates into and out of representative state (S_r, S_a, I_r, I_a) are also shown.

Table 2. State space of the PCMC in the left column and an NCMC in the right column for a 2-node N/VSIS under (A2).

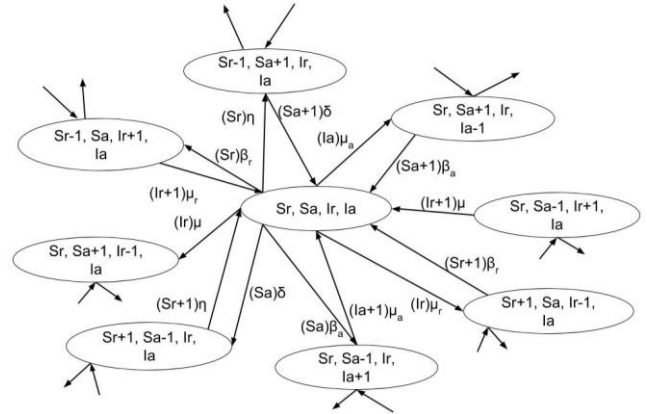| $x = (S_r, S_a, I_r, I_a), S_r, S_a, I_r, I_a \in \{0,1, 2\}, S_r + S_a + I_r + I_a = 2, \|S_{PC}\| = 10$	$x = (x_1, x_2), x_i \in \{S_r, S_a, I_r, I_a\}, i = 1,2, \|S_{NC}\| = 16$
$(S_r, S_a, I_r, I_a) = (0,0,0,2)$	$(x_1, x_2) = (I_a, I_a)$
$(S_r, S_a, I_r, I_a) = (0,0,1,1)$	$(x_1, x_2) = (I_r, I_a), (I_a, I_r)$
$(S_r, S_a, I_r, I_a) = (0,0,2,0)$	$(x_1, x_2) = (I_r, I_r)$
$(S_r, S_a, I_r, I_a) = (0,1,0,1)$	$(x_1, x_2) = (S_a, I_a), (I_a, S_a)$
$(S_r, S_a, I_r, I_a) = (0,1,1,0)$	$(x_1, x_2) = (S_a, I_r), (I_r, S_a)$
$(S_r, S_a, I_r, I_a) = (0,2,0,0)$	$(x_1, x_2) = (S_a, S_a)$
$(S_r, S_a, I_r, I_a) = (1,0,0,1)$	$(x_1, x_2) = (S_r, I_a), (I_a, S_r)$
$(S_r, S_a, I_r, I_a) = (1,0,1,0)$	$(x_1, x_2) = (S_r, I_r), (I_r, S_r)$
$(S_r, S_a, I_r, I_a) = (1,1,0,0)$	$(x_1, x_2) = (S_r, S_a), (S_a, S_r)$
$(S_r, S_a, I_r, I_a) = (2,0,0,0)$	$(x_1, x_2) = (S_r, S_r)$

Fig.1 Partial PCMC rate transition diagram with transitions into and out of state (S_r, S_a, I_r, I_a), with $0 < S_r, S_a, I_r, I_a < N$ and $S_r + S_a + I_r + I_a = N$.

The transient state probabilities for the PCMC can be solved from the forward Chapman-Kolmogorov equation [3].

$$d\vec{\pi}(t)/dt = \vec{\pi}(t)Q, \quad (6)$$

where $\vec{\pi}(t)$ is a row vector containing $|S_{PC}|$ state probabilities that form a distribution $\sum_{x \in S_{PC}} \pi_x(t) = 1$, and Q is an $|S_{PC}| \times |S_{PC}|$ transition rate matrix populated with transition rates. As an example, Q matrix of the 2-node PCMC is

$$Q = \begin{bmatrix} q_{1,1} & 0 & 0 & 2\mu_a & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & q_{2,2} & 0 & \mu & \mu_a & 0 & \mu_r & 0 & 0 & 0 \\ 0 & 0 & q_{3,3} & 0 & 2\mu & 0 & 0 & 2\mu_r & 0 & 0 \\ \beta_a & 0 & 0 & q_{4,4} & 0 & \mu_a & \delta & 0 & 0 & 0 \\ 0 & \beta_a & 0 & 0 & q_{5,5} & \mu & 0 & \delta & \mu_r & 0 \\ 0 & 0 & 0 & 2\beta_a & 0 & q_{6,6} & 0 & 0 & 2\delta & 0 \\ 0 & \beta_r & 0 & \eta & 0 & 0 & q_{7,7} & 0 & \mu_a & 0 \\ 0 & 0 & \beta_r & 0 & \eta & 0 & 0 & q_{8,8} & \mu & \mu_r \\ 0 & 0 & 0 & 0 & \beta_r & \eta & \beta_a & 0 & q_{9,9} & \delta \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\beta_r & 2\eta & q_{10,10} \end{bmatrix},$$

where $q_{i,i} = -\sum_{j \neq i} q_{i,j}, i = 1, \cdots, 10$. The unique equilibrium $\vec{\pi}(\infty)$ of the PCMC can be solved from a set of $|S_{PC}|$ independent linear equations among the $|S_{PC}| + 1$ in

$$\vec{\pi}(\infty)Q = \vec{0}, \ \sum_{x \in S_{PC}} \pi_x(\infty) = 1. \quad (7)$$

Remark 2. A PCMC and an NCMC built for the same network are equivalent, as long as (A2) is assumed for both. To see this through the simple example of the 2-node N/VSIS network, aggregate each pair of the NCMC states that appear in the same row in the right column of Table 2, and order the (aggregated) states as they appear in the table. Then the PCMC and NCMC represented by (6) have exactly the same Q matrix.

2.3 Evolution of mean populations

In this section, the PCMC built for the modified N/VSIS network is used to predict network-wide prevalence of CTIs. Denote by $(\pi_{S_r}, \pi_{S_a}, \pi_{I_r}, \pi_{I_a})(t)$ the set of marginal distributions of the PCMC

$$\pi_{S_r=i}(t) = \sum_{x,S_r=i} \pi_x(t), \qquad \pi_{S_a=i}(t) = \sum_{x,S_a=i} \pi_x(t),$$
$$\pi_{I_r=i}(t) = \sum_{x,I_r=i} \pi_x(t), \qquad \pi_{I_a=i}(t) = \sum_{x,I_a=i} \pi_x(t). \quad (8)$$

Components $\pi_x(t)$ of $\vec{\pi}(t)$ are solved from (6). Mean population evolution (or mean population dynamics) can then be found as follows.

$$E\{S_r(t)\} = \sum_{i=1}^{N} i \times \pi_{S_r=i}(t), E\{S_a(t)\} = \sum_{i=1}^{N} i \times \pi_{S_a=i}(t),$$

$$E\{I_r(t)\} = \sum_{i=1}^{N} i \times \pi_{I_r=i}(t), E\{I_a(t)\} = \sum_{i=1}^{N} i \times \pi_{I_a=i}(t). \quad (9)$$

Using (6), (8), and (9), the evolution of mean populations for the PCMC is computed for the modified N/VSIS network with $N = 19$ nodes. The four normalized mean population dynamics as functions of time are displayed in Fig. 2. Rates $(\eta, \delta, \beta_r, \beta_a, \mu_r, \mu_a, \mu)$ are set at $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{30}, \frac{1}{60}, \frac{20}{3}, \frac{80}{3}, \frac{20}{3}\right)$ (1/unit time). It is seen that all nodes start in population group S_r, and gradually spread into other three population groups. At steady state, the population is divided almost equally between the non-vigilant susceptible and vigilant susceptible populations. At no point during the evolution do the infected populations become significant. The long-run mean populations settle at $E\{S_r(\infty)\} = 8.9598$, $E\{S_a(\infty)\} = 10.0084$, $E\{I_r(\infty)\} = 0.0262$, and $E\{I_a(\infty)\} = 0.0056$. The observed dependence of the total infectious population on the parameters is as expected, increasing with respect to increasing δ, β_r and β_a, and decreasing with respect to increasing μ_r, μ_a, μ and η, while $\beta_r > \beta_a$ and $\mu_r < \mu_a$ distinguish the non-vigilant and vigilant populations.

Remark 3. We name parameters $\eta, \mu_r, \mu,$ and μ_a in Table 1 as controllable parameters, which can be adjusted to suppress the total infectious mean population $E\{I_r(\infty) + I_a(\infty)\}$ in the long run at a cost. This can be formulated as a Markov decision problem [1], [3], whose solutions will be reported separately in the near future.
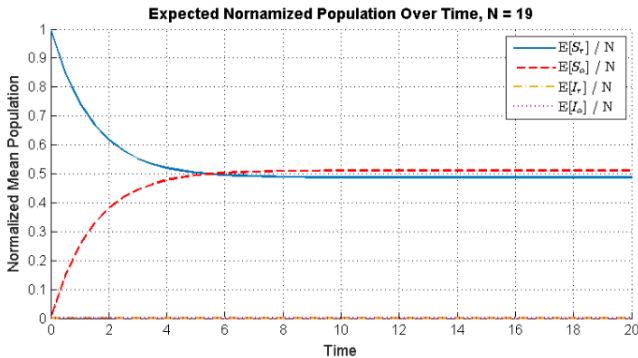

Fig.2 Mean populations solved using (6), (8), (9) for the PCMC of a 19-node N/VSIS network initialized at $\pi_{(19,0,0,0)}(0) = 1$.

The N/VSIS network size chosen is limited by the computational power required to solve (6), which is of dimension $|S_{PC}| = (N+3)!/N!/3! = 1540$ for a 19-node network. Note that the dimension of Equation (6) has been reduced from $|S_{NC}| = 4^{19} > 2.7 \times 10^{11} \gg 1.54 \times 10^3 = |S_{PC}|$ for a 19-node NCMC. To study the mean population evolution of PCMCs with more than a few tens of nodes, we propose to realize a PCMC as a queueing network in the next section.

3. MEAN POPULATION OF A PCMC SOLVED FROM A CLOSED QUEUEEING NETWORK REALIZATION

This section formalizes a closed queueing network realization of a PCMC for the modified N/VSIS network. A closed queueing network [3] is a discrete-state system of N entities, which move among a network of m queueing nodes to receive specific services. Such a queueing network can be simulated using a discrete-event simulation software package, such as SimEvents [8]. The discussion in this section also draws a numerical comparison between the PCMC of Section 2.2 and its queueing network realization for the same 19-node N/VSIS network, and for larger networks in terms of their efficiency in evaluating the evolution mean population evolution. Extension of queueing network realization to NCMC of more general networks is then briefly discussed.

3.1 Queueing network realization of a PCMC

The four queueing nodes (Q-nodes), respectively named as susceptible risk-seeking, susceptible risk-aversion, infectious risk-seeking, and infectious risk-aversion Q-nodes, are abbreviated as S_r, S_a, I_r, and I_a in Fig. 3. Each Q-node can house up to N entities in its N servers. Thus, each Q-node is of the M/M/N/N type, meaning Poisson arrivals / exponential service time distributions / N identical servers per Q-node / queue capacity equal to N at each Q-node. Each entity here corresponds to a computing node (C-node) in the N-node N/VSIS network. Service is provided as soon as an entity arrives at a Q-node, as there are only N entities (C-nodes) in the closed queueing network. Because symbols S_r, S_a, I_r, I_a are used to denote the names of the Q-nodes, we avoid using them to also define the state space for the above queueing network.

$$S_{QN} = \{(n_1, n_2, n_3, n_4) | 0 \leq n_1, n_2, n_3, n_4 \leq N,$$

$$n_1 + n_2 + n_3 + n_4 = N\}. \quad (10)$$

whose element $x = (n_1, n_2, n_3, n_4)$ is the composite queue length at the four queueing nodes named S_r, S_a, I_r, I_a, respectively, as depicted in Fig.3.

Theorem 1. The PCMC with state space defined in (4) constructed for the 4-population N-computing-node N/VSIS network in Section 2.2 fully coincides with the Markov chain with state space defined in (10) representing the 4-queueing node N-entity closed queueing network of Fig. 3, where the service rates and routing probabilities are specified.

Proof. Consider the state space defined in (10) for the closed queueing network in Fig.3. The entities in the queueing network have been equated to the C-nodes in the N/VSIS network, the queue lengths (n_1, n_2, n_3, n_4) in the four Q-nodes have been equated to the numbers of C-nodes in their respective four population groups (S_r, S_a, I_r, I_a). An entity arrival at a Q-node implies a transition of a C-node into the population group hosted by the corresponding Q-node, and the service time of an entity at a server equals to the holding time of a C-node in the corresponding population group.

Specific service rates are now assigned to the servers in the four Q-nodes. They are $\sigma_1 = \beta_r + \eta$ for any server in the susceptible risk-seeking Q-node, $\sigma_2 = \beta_a + \delta$ for any server in the susceptible risk-aversion Q-node, $\sigma_3 = \mu_r + \mu$ for any server in the infectious risk-seeking Q-node, and $\sigma_4 = \mu_a$ for any server in the infectious risk-aversion Q-node. Note that the inverse of the service rates are the average holding times of the C-nodes in their respective population groups.

A routing probability at the output of a Q-node indicates the likelihood of a departing entity's being routed to the input of another Q-node. They are $p_{12} = \frac{\eta}{\beta_r + \eta}$, $p_{13} = \frac{\beta_r}{\beta_r + \eta}$ from Q-node S_r to Q-nodes S_a, I_r, respectively; $p_{21} = \frac{\delta}{\beta_a + \delta}$,

$p_{24} = \frac{\beta_a}{\beta_a + \delta}$ from Q-node S_a to Q-nodes S_r, I_a, respectively; $p_{31} = \frac{\mu_r}{\mu_r + \mu}$, $p_{32} = \frac{\mu}{\mu_r + \mu}$ from Q-node S_a to Q-nodes S_r, S_a, respectively; and $p_{42} = 1$ from Q-node I_a to Q-node S_a. These correspond to event probabilities at a given PCMC state.

Since state space S_{QN} defined in (10) for the queueing network is the same as state space S_{PC} defined in (4) for the corresponding PCMC, the construction of the Markov chain for the queueing network depicted in Fig.3 generates the same rate transition diagram shown in Fig.1, which uniquely defines transition rate matrix Q in (6). More specifically, the probability flow balance equation [3] of the Markov chain of the queueing network governing the evolution of state probability $\pi_{(n_1, n_2, n_3, n_4)}$ at composite queue length (n_1, n_2, n_3, n_4) in the four Q-nodes S_r, S_a, I_r, I_a respectively is given by

$$\dot{\pi}_{(n_1,n_2,n_3,n_4)} = - \sum_{i:n_i > 0} \sigma_i \pi_{(n_1,n_2,n_3,n_4)}$$
$$+ \sum_{j:n_j > 0} \sum_i p_{ji}\, \sigma_j \pi_{(\cdots,n_j+1,\cdots,n_i-1,\cdots,)},$$

where $0 \le n_1, n_2, n_3, n_4 \le N$ and $\sum_{i=1}^4 n_i = N$. This gives rise to $|S_{QN}| = (N+3)!/(N! \times 3!)$ differential equations on state probabilities matching those $|S_{PC}|$ differential equations in (6). Therefore, queue length distribution $\{\pi_{(n_1,n_2,n_3,n_4)}\}$ for the 4-queueing-node N-entity queueing network in Fig.3 and population distribution $\{\pi_{(S_r,S_a,I_r,I_a)}\}$ for the 4-population N-computing-node PCMC are identical. Using (8) and (9), the mean evolution of the queue lengths evaluates the mean evolution of the PCMC populations.
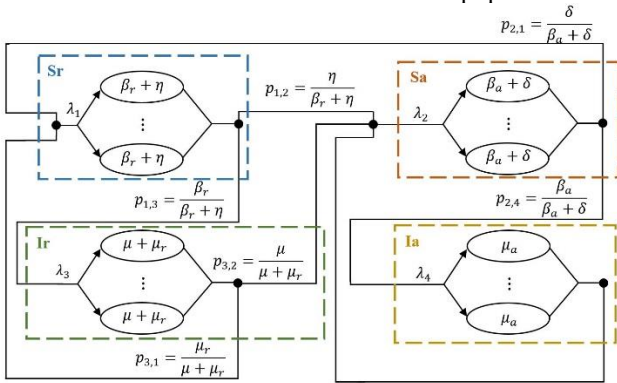

Fig.3 A closed queueing network realization of a PCMC for a 4-population N-computing-node (C-node) N/VSIS network. Each queueing node (Q-node) is of the M/M/N/N type. Routing probabilities at the output of each Q-node indicates the likelihood of a departing C-node (entity) from a population group (Q-node) being routed to the input of another Q-node

Fig.4 shows the evolution of mean queue lengths in the four queueing nodes of the 19-entity queueing network of Fig.3. Each sample path in the figure is the pointwise average of 20 independent replications generated by SimEvents [8] initiated at state $(S_r, S_a, I_r, I_a) = (19,0,0,0)$. The PCMC rate parameters $(\eta, \delta, \beta_r, \beta_a, \mu_r, \mu_a, \mu)$ used in Section 2.3 are retained, which now enter the service rates and routing probabilities in the queueing network in Fig.3. They define distributions from which random variates of entity service times are sampled.
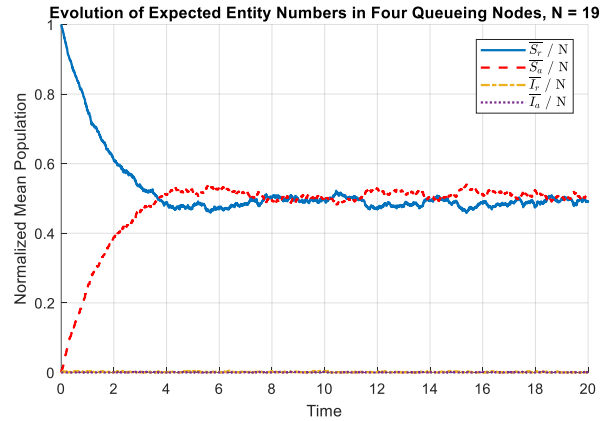

Fig.4 Evolution of expected queue lengths in the 4-queueing node 19-entity queueing network in Fig.3, generated by SimEvents that averages 20 independent replications initialized at (n_1, n_2, n_3, n_4) =(19,0,0,0).

With the equivalence between a PCMC and its queueing network realization established and numerically verified as seen in Fig.2 and Fig.4, we now move to examine the mean population evolution of a 1500-node N/VSIS network using discrete event simulation of a 1500-entity queueing network realization of its PCMC. Fig.5 shows the SimEvents generated average sample paths of 20 independent replications on the evolution of mean queue lengths using the same set of rate values. It is interesting to note that how the mean queue lengths (mean populations) evolve in time remains unchanged for this N/VSIS network of nearly 100 times larger.
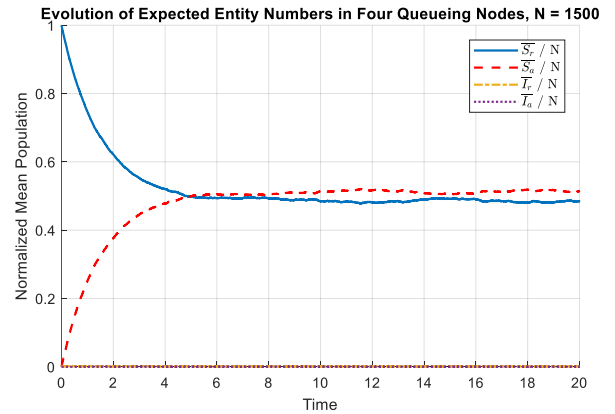

Fig.5 Evolution of expected queue lengths in the 4-Q-node 1500-entity queueing network in Fig. 3 initialized at (1500,0,0,0) with 20 independent replications.

To cover a wider range of the N/VSIS network size, PCMCs realized as closed queueing networks are programmed in MATLAB for evaluation of the evolution of the four mean queue lengths. As the N/VSIS network size increases from 15 to 1,500 to 150,000 nodes, the CPU time increases from 0.33 to 29.3 to 2943.8 seconds. This is a linear growth of CPU time. Each evaluation involves 10 replications of 20-second long simulation time, and is carried out by a 2.6 GHz Intel Xeon processor with 8GB RAM.

Remark 4. The queueing network realization of a PCMC enables fully rigorous evaluation of the mean populations without resorting to any mean field approximations and without having to solve for (6). The computational benefit

is gained through analysing the statistics of a few quantities easily observable from discrete event simulations and directly linked to the performance measures sought. More specifically, only the queue length profile samples at the four Q-nodes are acquired and averaged over a sufficient number of independent replications to estimate the mean population profiles of the PCMC. The samples are readily collected from the simulated queueing network in Fig.3 with a defined set of Poisson clock structure and routing probabilities mapped from the PCMC. On the other hand, seeking mean population dynamics via evaluating (6), (8) and (9) for a PCMC requires at least multiples of N^6 floating point operations [2] in each discrete time step, which worsens to an exponential function of N per time step for an NCMC.

3.2 Some Extensions

Some immediate extensions of the above work are now briefly discussed.

(i) Queueing network realization of an m-population PCMC. Equivalence between the PCMC of an m-population N-node computer network and the Markov chain of an m-queueing-node N-entity queueing network can be readily established for any m population groups following the same argument as in the proof of Theorem 1, which has been established for $m = 4$.

(ii) Queueing network realization of an NCMC modelling an arbitrary directed network of heterogeneous nodes. For simplicity, we return to the example of a 4-population N-computing-node N/VSIS network, for which, however, the assumption of homogeneous nodes and node interactions is relaxed. Now the servers in each Q-node are ordered according to the cardinal numbers assigned to the C-nodes. Sever $i, i = 1, \cdots, N$, in Fig.3 can only host C-node i, with service rate $\beta_{r,i} + \eta_i$, or $\beta_{a,i} + \delta_i$, or $\mu_{r,i} + \mu_i$, or $\mu_{a,i}$ in Q-node S_r, or S_a, or I_r, or I_a, respectively. As a result, at most one of the ith servers in the four Q-nodes can be occupied by C-node i at a time. The routing probabilities following each Q-node in Fig.3 must be modified to have subscript i as well, and thus its value becomes departure entity-dependent. In this setting, the queueing network state space S_{QN} becomes identical to state space S_{NC} of an NCMC. An example of naming the NCMC states is given in Table 2 when N=2. Based on the argument in Remark 4, significant computational advantage can be expected in the evaluation of the NCMC mean population dynamics with a queueing network realization.

4. CONCLUSIONS

A strongly connected, homogeneous N/VSIS network [5] was placed in the class of node-centric networked Markov chains (NCMC) whose mean-field-like approximations conform to the traditional epidemiological models. To study the evolution of mean infectious populations of NCMCs for large networks, a population-centric Markov chain (PCMC) with significantly reduced state space size was proposed. Computational complexity for evaluation of mean population dynamics was further reduced to linear growth with respect to network size N by a proposed queueing network realization of the PCMC. Two extensions were discussed: (i) queueing network realization of PCMCs for m-population networks (m ≠ 4), and (ii) queueing network realization of NCMCs as an alternative computational framework for evaluation of mean population dynamics of networks with heterogeneous computing nodes.

Many challenges remain. We seek in the future to design cost constrained policies to control the global prevalence of CTIs under the proposed new modelling framework (Remark 3). We look for ground-truth data to estimate the rate parameters [4] in the structured PCMCs for specific networks using labelled attack data (Remark 1). We also seek to formally establish the NCMC-queueing network equivalence, briefly discussed in Section 3.2 (ii), for heterogeneous networks and quantify the expected complexity reduction afforded by the queueing network realization (Remark 4).

REFERENCES

[1] Bertsekas, D., *Dynamic Programming and Optimal Control, Vols. I and II*, Athena Scientific, 2005.

[2] Boyd, S. and L. Vandenberghe, *Convex Optimization,* Cambridge University Press, 2004.

[3] Cassandras, C. G., and S. Lafortune, *Introduction to Discrete Event Systems*. New York, NY: Springer Science+Business Media, 2008.

[4] Kay, Steven M., *Fundamentals of statistical signal processing, volume I: Estimation Theory*, Prentice Hall, 1993.

[5] Kelley, T., and L.J. Camp, Online Promiscuity: Prophylactic Patching and the Spread of Computer Transmitted Infections, Workshop on the Economics of Information Security (WEIS) 2012, June 25-26, Berlin, Germany, 2012.

[6] Kermack, W.O., and McKendrick, A. G., A Contribution to the Mathematical Theory of Epidemics, *Proceedings of the Royal Society of London. Series A, Mathematical and Engineering Sciences*, vol. 115, pp. 700-721, 1927.

[7] Kephart, J., and White, S. Measuring and modeling computer virus prevalence. Proceedings 1993, *IEEE Computer Society Symposium on Research in Security and Privacy*, pp.2-15, 1993.

[8] MathWorks, *SimEvents User's Guide*, The MathWorks Inc., 2018.

[9] Nowzari, C., V.M. Preciado, G.J. Pappas. Optimal resource allocation for control of networked epidemic models. *IEEE Transactions on. Control of Network. Systems,* vol.4, pp,159-169, 2017.

[10] Paarporn, K., C. Eksin, J. S. Weitz, and J. S. Shamma. Networked SIS Epidemics with Awareness. *IEEE Transactions on Computational Social Systems*, vol. 4(3), pp. 93-103, 2016.

[11] Pastor-Satorras, R. , C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, *Reviews of Modern Physics*, vol. 87, pp. 925–979, 2015.

[12] Ruhi, N.A., and B. Hassibi, SIRS Epidemics on Complex Networks: Concurrence of Exact Markov Chain and Approximated Models, *Proceedings of IEEE Conference on Decision and Control*, pp. 2919-2926, 2015.

[13] Shakeri, H., F. D. Sahneh, C. Scoglio, P Poggi-Corradini, and V. M. Preciado. Optimal information dissemination strategy to promote preventive behaviors in multilayer epidemic networks. *Mathematical Biosciences and Engineering*, vol. 12(3), pp.609–623, 2015.

[14] Van Mieghem, P., J. Omic, and R. Kooij, Virus spread in networks, IEEE/ACM Transactions on Networking, vol. 17, pp. 1–14, 2009.

[15] Wang, Y., and Wang, C., Modeling the effects of timing parameters on virus propagation. *Proceedings of the 2003 ACM workshop on Rapid Malcode - WORM'03*, ACM Press, p. 61, 2013.