Optimal Caching Time for Epidemic Content Dissemination in Mobile Social Networks

Hsiang Hsu, and Kwang-Cheng Chen, *Fellow, IEEE*

***Abstract*—To facilitate content distribution and diffusion, efficacious caching strategy plays an important role in content dissemination control, especially in dynamic mobile social networks (MSNs), where contents spreading and accessing rely mainly on opportunistic contacts in physical proximity. Since content dissemination much resembles epidemic dynamics, two caching control schemes, caching at external BS and cooperative in-network caching, are investigated to assay the system behaviours and performance via epidemic dynamics. When time dynamic is considered, we provide a more realistic scenario where the cost of caching is related to the time duration of caching; hence optimal control theory are exploited to determine the optimal caching time for the content spreading. Moreover, we provide proactive caching analysis as a preventive system response to handle severe outbreak of the epidemic content, which would often cause instantaneous service burden in the system. Finally, virality is shown to be an important content feature when implementing caching. This research, from the aspect of system dynamics, paves novel avenues to content dissemination and caching utilization in mobile social networks.**

***Index Terms*—Caching, content dissemination, epidemic model, mobile social networks, optimal control, resource utilization.**

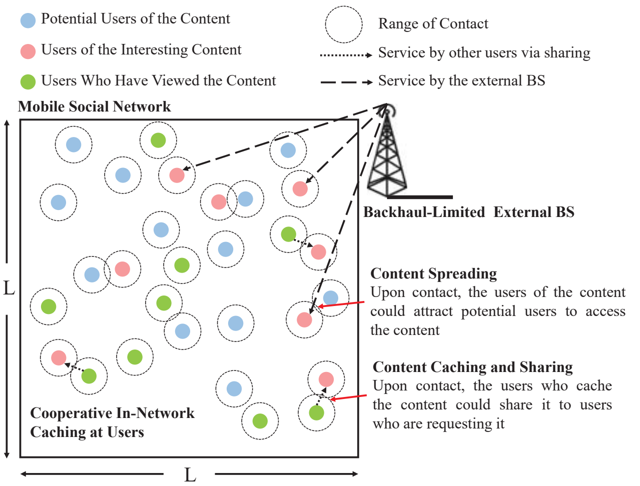

Fig. 1: The mechanisms of content spreading and caching in mobile social networks in a snapshot. The users of the content could attract potential users to access it upon contacts within the range. The content can be accessed by interested users in two ways: (1) By an backhaul-limited external BS, which service rate would increase if it caches the content (section II-A). (2) By sharing from cooperative users who cache the content after accessed it in the MSN upon contacts (section II-B).

I. Introduction

With the advent of mobile communication technologies, manifold mobile social networks (MSNs) have emerged, facilitating direct exchange of contents among mobile users through large complex mechanism of contacts when they are in physical proximity, including news, comments, and other mobile social media [1]. In particular, the direct exchange among users forms a new spreading behavior and thus dynamic of contents. That is, in addition to traditional social media distributed to users through mobile infrastructure networks and cloud servers (*e.g.* a base station, BS), users of an interesting content could attract users of potential interests to access such content through social interactions [2]. This process is exactly analogous to the well-known epidemics process, and the users of such content can be viewed as infected users (Fig. 1). Therefore, due to the fact that this content dissemination process much resembles the spread of epidemics, epidemic modelling [3] has been used in myriads of areas to investigate the information dynamics in MSNs [4]–[7].

As efficient content access and dissemination is the key to modern mobile networks, proper content caching [8]–[10] has been recently suggested as an effective methodology in MSNs to address the booming traffic volume of social applications [11]. When caching mechanism is imposed in the MSN, the users who have viewed the content could store it, thereby sharing the content to others who also request it in physical proximity (Fig. 1). Currently, the investigation of caching focuses on the inhomogeneous popularities of contents to develop caching mechanisms. However, in MSNs, where the information dynamics becomes a practical aspect, popularities could only depict static characteristics of social contents; alternatively, the virality of the content [12] turns out to be a more realistic feature of social contents, and more adequate for epidemic model. Moreover, a further practical engineering consideration of system cost of caching, time duration of caching, is often neglected; however, since caching a content usually occupied the limited storage capacity, the time duration of caching a content should be taken into account as the metric of cost. Consequently, in time dynamic systems like MSNs, the time instant for caching a content must compromises between the cost of caching, and the system dynamics. To be more specific, early caching leads to massive waste of storage capacity, whereas late caching reduces efficiency of caching and sharing for content dissemination among users. To the best of our knowledge, the optimal caching time for epidemic contents still remains an open but critical technical issue in this complex dynamic system. To identify optimal time for caching is even more complicated in mobile social networks, owing to

H. Hsu and K.-C. Chen are with the Graduate Institute of Communication Engineering, National Taiwan University, Taipei 10617, Taiwan. (Email: r03942046@ntu.edu.tw, ckc@ntu.edu.tw)

This research was supported in part by the U.S. Air Force under Grant AOARD-14-4053, and in part by the Ministry of Science and Technology of Taiwan under Grant MOST 104-2622-8-002-002, and sponsored by MediaTek Inc., Hsin-chu, Taiwan.

978-1-4799-6664-6/16/$31.00 ©2016 IEEE

the fact that mobility nurtures the spread of epidemic contents and makes precise analysis difficult.

Hence, to carry out the analysis on this dynamic system, in this paper, we relate the content dissemination in MSNs to an epidemic, and we consider a more general scenario that caching is another kind of resource to help spreading the content. Incorporating the spreading of content and effect of caching, we formulate the state evolutionary equations of the system with two cases. First, we consider a BS external to the MSN which serves the users with the content, referring to a traditional global serving mechanism in section II-A. It is realistic that the service rate of the BS will increase if the content is cached at the BS, since the deficiency of backhaul capacity is generally a bottleneck in serving the mobile users [13]. Second, we consider in-network caching at mobile users in the MSN as a cooperative mechanism to help distribute the content in section II-B. The system performance is to minimize the number of unserved users in the observed time and the cost of caching as well. Leveraging optimal control theory [14], the optimal caching time is derived in section III. Moreover, we also explore proactive caching analysis to handle the outbreak of the epidemic content in the early stage as a common phenomenon in social networks, providing preventive system response at the imminent phase (section IV). The evaluation of system performance and the optimal caching time are provided in section V, together with the effect of virality of the content. Finally, section VI concludes this paper.

II. System Model

In order to analyze the complex and opportunistic mechanism of contacts and the time dynamic relationship among users in the MSN, we adopt the non-linear ordinary differential equations (ODEs) of epidemic model [3] to describe the system. Especially, the susceptible-infected-recovered (SIR) model is utilized as the state evolutionary equations for control process [15]. To bridge the analogue between the SIR model and composition of users in the MSN, we denote three compartments $S(t)$, $I(t)$, and $R(t)$ as the susceptible, infected, and recovered population of users inside the MSN at time t. Therefore, in the MSN, the susceptible users in $S(t)$ represent the potential viewers of the content; the infected users in $I(t)$ are contagious users who want but still waiting for the content; the recovered user in $R(t)$ stands for those who have accessed (have been served with) the content and will not request it again for good. An infected user could attract a susceptible user to access the content when they make contacts within a range of contact δ. Furthermore, N denotes the total population in this network, *i.e.* $S(t) + I(t) + R(t) = N$, $\forall t \geq 0$. In the MSN, the N mobile users move randomly in a $L \times L$ square area, in which the content spreads and users interact among each other. In the subsequent paragraphs we formulate the state equations of *caching at external BS* and *cooperative in-network caching* schemes.

A. Caching at External BS

In this scheme, infected users could only access the content from a BS external to the MSN (*e.g.* a femto-cell BS) with limited backhaul, as a traditional means to access the content. The limited backhaul resource becomes a bottleneck for the service rate $u(t)$ of the BS. However, when caching the content at BS, the service rate will increase due to the mitigation of traffic load on backhaul by caching [13]. The service rate thus can be expressed as

$$u(t) = \begin{cases} \kappa_1, & t < T_C \\ \kappa_2, & t \geq T_C \end{cases} \tag{1}$$

T_C is the time to cache the content at the BS, and $0 \leq \kappa_1 \leq \kappa_2 \leq 1$. κ_2 and κ_1 are the service rates with/ without caching respectively.

The variation of infected population is determined by the pairwise virality of the content, rate of encountered users and the fraction of susceptible users to the epidemic content. The recovered population is controlled by the service rate of the BS. Putting all together, the state equations can be formulated as

$$\begin{cases} \dot{I}(t) = \nu\eta\hat{S}(t)I(t) - u(t)I(t), \\ \dot{R}(t) = u(t)I(t), \\ \dot{S}(t) + \dot{I}(t) + \dot{R}(t) = 0, \end{cases} \tag{2}$$

where $\hat{S}(t) = S(t)/N$ is the normalized susceptible population at time t; $\nu \in \mathbb{R}$ is the virality of the content, which can be interpreted as the penchant for susceptible users to request the content upon contact with infected users in the range δ; $\eta = \pi\delta^2/L^2$ is the average number of encountered users per unit time. The last equation is implied from fixed total population.

B. Cooperative In-Network Caching

In the previous scheme, the feature of the MSN, direct opportunistic exchange of contents, is utilized for content dissemination. However, this feature could be further exploited when caching is imposed on users in the MSN. In other words, the recovered users could cache the epidemic content and share it to socially satisfy other users, which can be treated as another means to serve users in mobile networks. This new serving mechanism, in fact, is a generalization of device-to-device (D2D) communication when we consider caching utilization in mobile networks [13]. The population of recovered users is than further increased by the cooperative sharing among users. The state equations can thus be formulated as

$$\begin{cases} \dot{I}(t) = \nu\eta\hat{S}(t)I(t) - u(t)I(t) - \phi(t)\eta\hat{R}(t)I(t), \\ \dot{R}(t) = u(t)I(t) + \phi(t)\eta\hat{R}(t)I(t), \\ \dot{S}(t) + \dot{I}(t) + \dot{R}(t) = 0, \end{cases} \tag{3}$$

where $\hat{R}(t) = R(t)/N$ is the normalized recovered population at time t,

$$\phi(t) = \begin{cases} 0, & t < T_C \\ \kappa, & t \geq T_C \end{cases} \tag{4}$$

and $\kappa \in [0, 1]$ is the socially cooperative sharing coefficient which can be interpreted as the pairwise willingness to share the epidemic content. Initially, the recovered users do not know whether to cache the content, and start to cache the content and provide auxiliary sharing of the content after time T_C. A direct observation from (3) shows that in-network caching proliferates the recovered population and therefore alleviates the growth of infected population. Moreover, comparing (2) and (3), caching at external BS is actually a degenerate case of cooperative in-network caching when $\kappa = 0$, which means that the total service of the content comes from the BS, as the traditional means.

III. Optimal Control and Caching Time

Since caching at external BS is a special case of in-network caching scheme when there is no caching at users and hence no direct content sharing, for the general case, we use the state equations of cooperative in-network caching in (3) to obtain the optimal control $u^*(t)$ of $u(t)$. As explained in section I, the cost of caching is related to the time duration of caching, since in the duration, the storage capacity for caching the content could not store other contents. Therefore, the cost of caching could be formulated into the cost of control $u(t)$ in the duration, since caching leads to higher service rate. Besides the cost, the efficient control of caching also depends on the number of recovered users. In more details, there is no recovered user in the MSN at the beginning, causing scarce sharing and inefficient caching; however, the number of recovered users will increase by time and hence enhances the sharing, thus rendering the time for caching a critical issue. Aiming to determine the optimal control, we exploit optimal control theory [14] to leverage the cost of caching and the efficiency of epidemic content dissemination, *i.e.* to minimize the cumulative number of users who want but still waiting for the content. Consequently, the goal is to minimize the performance measure functional J, where

$$J = \int_{T_i}^{T_f} I(t)^{\beta} + \frac{1}{\alpha} u(t)^{\alpha} dt \tag{5}$$

and $u(t)$ taking its α-power form, $\alpha \geq 0$; higher α means larger cost of caching. $\beta \geq 0$ represents the requirement of system dynamic; larger β and hence larger cost $I(t)^{\beta}$ indicates that we desire fewer unserved users in the system. T_f is the completion time for observation which is assumed to be free and T_i is the initial time set to be 0.

In optimal control theory, we construct the functional Hamiltonian $\mathcal{H}$ by applying Euler-Lagrange equation as

$$\begin{aligned}
&\mathcal{H}\left(I(t), R(t), u(t), \Lambda_I(t), \Lambda_R(t)\right) \\
&= I(t)^{\beta} + \frac{1}{\alpha} u(t)^{\alpha} \\
&+ \Lambda_I(t)\left[\nu\eta\hat{S}(t)I(t) - u(t)I(t) - \phi(t)\eta\hat{R}(t)I(t)\right] \\
&+ \Lambda_R(t)\left[u(t)I(t) + \phi(t)\eta\hat{R}(t)I(t)\right]
\end{aligned} \tag{6}$$

from which the costate variables $\Lambda_I^*(t)$ and $\Lambda_R^*(t)$ are the partial derivatives of the Hamiltonian with respect to $I(t)$ and $R(t)$:

$$\begin{aligned}
\dot{\Lambda}_I^*(t) &= -\frac{\partial \mathcal{H}}{\partial I} = -\beta I(t)^{\beta-1} \\
&- \Lambda_I^*(t)\left[\nu\eta\frac{N - 2I(t) - R(t)}{N} - u(t) - \phi(t)\eta\frac{R(t)}{N}\right] \\
&- \Lambda_R^*(t)\left[u(t) + \phi(t)\eta\frac{R(t)}{N}\right]
\end{aligned} \tag{7}$$

$$\begin{aligned}
\dot{\Lambda}_R^*(t) &= -\frac{\partial \mathcal{H}}{\partial R} \\
&= \Lambda_I^*(t)\left[\nu\eta\frac{I(t)}{N} + \phi(t)\eta\frac{I(t)}{N}\right] - \Lambda_R^*(t)\phi(t)\eta\frac{I(t)}{N}
\end{aligned} \tag{8}$$

with boundary conditions $\dot{\Lambda}_I^*(T_f) = \dot{\Lambda}_R^*(T_f) = 0$. Assuming that all of the state and costate variables are according to their values for the optimal control $u^*(t)$, we rewrite the Hamiltonian in (6) with the switching function

$$\begin{aligned}
\theta^*(t) &:= \Lambda_I^*(t)I(t) - \Lambda_R^*(t)I(t) \\
&= \left[\Lambda_I^*(t) - \Lambda_R^*(t)\right] I(t)
\end{aligned} \tag{9}$$

yielding

$$\begin{aligned}
&\mathcal{H}\left(I^*(t), R^*(t), u(t), \Lambda_I^*(t), \Lambda_R^*(t)\right) \\
&= I(t)^{\beta} + \frac{1}{\alpha} u(t)^{\alpha} - \theta^*(t)u(t) \\
&+ \eta I^*(t)\left[\Lambda_I^*(t)\left(\nu\hat{S}(t) - \phi(t)\hat{R}(t)\right) + \Lambda_R^*(t)\phi(t)\hat{R}(t)\right]
\end{aligned} \tag{10}$$

By Pontryagin's minimum principle [16], the unconstrained optimal control $U^*(t)$ with free end time T_f that minimizes the performance measure J is the solution of the equation $\frac{\partial \mathcal{H}}{\partial u} = 0$. From the reformed Hamiltonian in (10), we have

$$U^*(t) = \theta^*(t)^{\frac{1}{\alpha-1}} \tag{11}$$

That is, $U^*(t)$ can be obtained by solving the state and costate variables in (3), (7) and (8). Moreover, with the acceptable control $u(t) \in [0, 1]$, the induced constrained optimal control $u^*(t)$ is

$$u^*(t) = \begin{cases} 0 & \text{if } \theta^*(t) \leq 0 \\ \theta^*(t)^{\frac{1}{\alpha-1}} & \text{if } \theta^*(t) \in (0, 1) \\ 1 & \text{if } \theta^*(t) \geq 1 \end{cases} \tag{12}$$

where the induced $u^*(t)$ is actually the saturated unconstrained optimal control.

Manifestly, a discrepancy exists between the service rate in (1) and the constrained optimal control $u^*(t)$, since in our derivative, $u^*(t)$ is acquired by presuming that immediate caching is permissible from initial time 0, whereas in (1), we consider a more realistic situation where the service rate increases only after the caching time T_C since it relieves the backhaul traffic load of the BS. Reflecting pragmatic implementation, the time T_C^* is the optimal caching time for caching at external BS and cooperative in-network caching that minimizes the performance measure J, *viz*

$$T_C^* = \underset{T_C}{\operatorname{argmin}}\, J \tag{13}$$

With this T_C^*, we can bridge the theoretical gap between (1) and (12). We will show that the solution in (13) provides an efficacious strategy for caching at external BS and in-network caching of the epidemic content in section V.

IV. Proactive Caching Analysis

In addition to optimal control approach and optimal caching time obtained in section III, we furnish proactive caching analysis, with an aim to determine optimal caching time in early stage to handle the outbreak of the epidemic content. The importance of proactive caching analysis lies on the fact that a severe outbreak of epidemic contents usually leads to suddenly massive requests beyond the system load, and it might be too late to cache the content at that time moment. Thus, analysis in the early-stage of the epidemic process helps store the content to handle the outbreak timely and efficiently for caching utilization. In the early stage, the state evolutionary equations (3) at time $\tau \ll T_f$ approximates a standard coupling regulator problem [14] when $S(\tau) \approx N$ and $I(\tau) \approx I_0 \ll N$:

$$\begin{cases} R(\tau) & = \exp\left(\frac{1}{N}\phi(t)\eta\tau I_0\right) - 1 \approx \frac{1}{N}\phi(t)\eta\tau I_0 \\ \dot{I}(t) & = \left[\nu\eta - u(t) - \left(\frac{1}{N}\phi(t)\eta\right)^2 \tau I_0\right] I(t) \end{cases} \tag{14}$$

By (1) and (4), we obtain

$$\begin{aligned} I(t) = &\quad I_0 \exp\left(\left(\nu\eta - u(t) - \left(\frac{1}{N}\phi(t)\eta\right)^2 \tau I_0\right) t\right) \\ = &\quad \begin{cases} I_0 \exp\left((\nu\eta - \kappa_1)\, t\right), & t < T_C \\ I_0 \exp\left(\left(\nu\eta - \kappa_2 - \frac{\kappa^2\eta^2}{N^2}\tau I_0\right) t\right), & t \geq T_C \end{cases} \end{aligned} \tag{15}$$

where I_0 is the initial infected population. The analytical form of the functional performance measure J in (5) becomes merely a function of T_C:

$$\begin{aligned} \widetilde{J} &= \int_0^{T_f} I(t)^\beta + \frac{1}{\alpha} u(t)^\alpha dt \\ &= \int_0^{T_C} I(t)^\beta + \frac{1}{\alpha}\kappa_1^\alpha dt + \int_{T_C}^{T_f} I(t)^\beta + \frac{1}{\alpha}\kappa_2^\alpha dt \\ &= \frac{I_0^\beta}{\beta A_1}\left[\exp\left(\beta A_1 T_C\right) - 1\right] + \frac{1}{\alpha}\kappa_1^\alpha T_C \\ &+ \frac{I_0^\beta}{\beta A_2}\left[\exp\left(\beta A_2 T_f\right) - \exp\left(\beta A_2 T_C\right)\right] + \frac{1}{\alpha}\kappa_2^\alpha \left(T_f - T_C\right) \end{aligned} \tag{16}$$

where $A_1 = \nu\eta - \kappa_1$ and $A_2 = \nu\eta - \kappa_2 - \frac{\kappa^2\eta^2}{N^2}\tau I_0$, for simplicity. Similar to optimal control, the optimal caching time when adopting proactive caching analysis can therefore be gained by

$$\widetilde{T_C^*} = \underset{T_C}{\operatorname{argmin}}\, \widetilde{J}. \tag{17}$$

The advantage of above proactive caching analysis is to convert the original functional J into a relatively simple function of time $\widetilde{J}$, via relaxing the computation to solve the non-linear ODEs in (3) for optimal control in section III. Nevertheless, it may suffer from potential waste of caching resources by overestimating the level of outbreak of the epidemic content, as a result of the assumption $S(\tau) \approx N$.

V. Performance Evaluation

We illustrate the system dynamics and the optimal caching time under different settings of the cost of caching α and the requirement of system dynamic β. To be more specific, α could be understood as the scarcity of caching resource in the network. In system view, a larger amount of contents circulating in the MSN sharing the storage at the BS makes storage capacity more scarce resources, and thus larger α. Meanwhile, β represents the requirement of quality of system dynamics. When the quality is required high, the number of users who want but are waiting for the content should be as few as possible, and β becomes large. Moreover, the optimal caching time is also investigated with respect to the virality of the content, as a different system feature than popularities of the contents. Pertaining to the simulation setup, N nodes are travelling in a square area with wrap-around condition via Lèvy walk mobility model to account more for human behavior [17], where the step size and pause time are accounted by a power-law distribution with negative exponent. We set the step size exponent to be 1.5 and the pause time exponent to be 1.38, respectively, which fits the real trace-based data collected in [18].

The system dynamics under caching at external BS scheme is shown in Fig. 2. The difference between the system dynamics of optimal control $u^*(t)$ and the simulation results adopting optimal caching time T_C^* results from the gap between the optimal time (1) and the optimal control (12) as discussed in section III. The solution of T_C^* in (13) is shown to be efficacious for epidemic content dissemination since it minimizes the performance measure J which is related to the number of users still waiting for the content in the observing area up to T_f. The solution of $\widetilde{T_C^*}$ from proactive caching analysis has a resembling effect on epidemic content dissemination, however, with a smoother decaying as a result of the approximation in section IV. Moreover, For in-network caching scheme, the system dynamic of users waiting for the content is further suppressed as shown in Fig. 3. For small κ (≈ 0.1), the impact of in-network caching is apparent, suggesting a slight incentive to promote social cooperation among users to share the content could result in improvements in system dynamics, and is an important design consideration in mobile social networks.

Subsequently, we investigate the optimal caching time with respect to system performance requirements, the cost of caching α and the requirement of system dynamic in the network β. In Fig. 4, the optimal caching time increases with the increase of α, since if there are many contents circulating in the MSN, a longer time is needed to decide whether to cache the viral content, in order to optimize the usage of storage capacity at the BS and the users. Nevertheless, since the number of users who have viewed the content grows by time, late caching time implies better chance to efficaciously utilize caching among the users for sharing. However, as we require the number of users waiting for the content as few as possible, the optimal caching time becomes early to handle the requirement. These two factors form a trade-off in design optimal caching utilization in mobile social networks to serve epidemic contents. Furthermore, we observe that in-network caching contributes to early caching by starting storing in recovered users to create more in-network sharers of the epidemic content. Similar effects for cost of caching and service requirements in the network can be found for proactive caching analysis in Fig. 5. It is worth mentioning that when the cost of caching is small (less than 1), the assumption that $S(t) \approx N$ and the low cost of caching lead to immediate caching than the results in Fig. 4, yet when the cost is large, it causes pessimistic outcome of late caching time than that in Fig. 4. In other words,

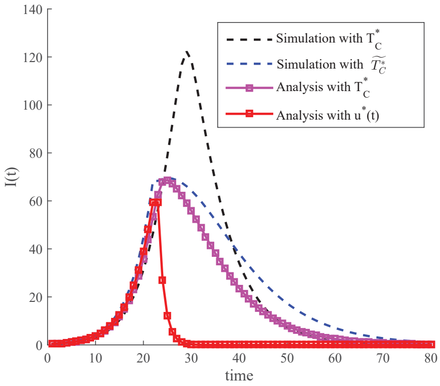

Fig. 2: System dynamics under caching at external BS. The difference between the information dynamic of optimal control $u^*(t)$ and the simulation results adopting optimal caching time T_C^* results from the gap between the optimal time (1) and the optimal control (12). $N = 1000$, $I_0 = 1$, $L = 100$, $\delta = 1$, $\nu = 1$, $\kappa_1 = 0.1$, $\kappa_2 = 0.2$, $\alpha = \beta = 2$, $\tau = 1$, $T_i = 0$, $T_f = 80$, $\Lambda_I(0) = 20$, $\Lambda_R(0) = 10$, with 1000 times of simulations

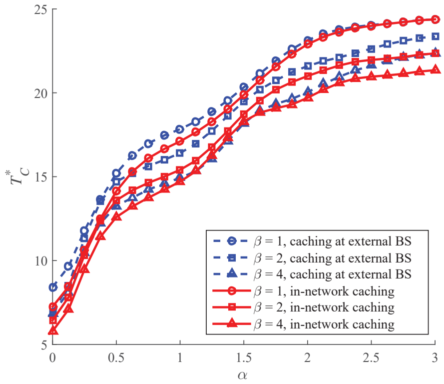

Fig. 4: Optimal caching time T_C^* under different (α, β) configurations, where α stands for the cost of caching and β stands for the requirement of system dynamic. Small cost of caching and high requirement results in immediate caching. $N = 1000$, $I_0 = 1$, $L = 100$, $\delta = 1$, $\nu = 1$, $\kappa_1 = 0.1$, $\kappa_2 = 0.2$, $\kappa = 0.15$, $T_i = 0$, $T_f = 150$, $\Lambda_I(0) = 20$, $\Lambda_R(0) = 10$.

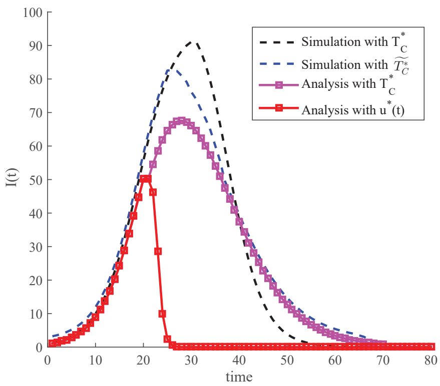

Fig. 3: System dynamics under in-network caching. With cooperative caching, the population of users waiting for the content is obviously reduced comparing to Fig. 2, where only caching at external BS is utilized. $N = 1000$, $I_0 = 1$, $L = 100$, $\delta = 1$, $\nu = 1$, $\kappa_1 = 0.1$, $\kappa_2 = 0.2$, $\kappa = 0.1$, $\alpha = \beta = 2$, $\tau = 1$, $T_i = 0$, $T_f = 80$, $\Lambda_I(0) = 20$, $\Lambda_R(0) = 10$, with 1000 times of simulations

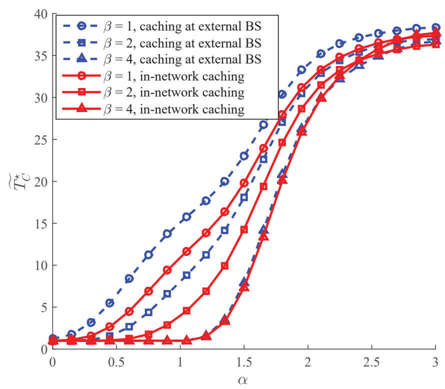

Fig. 5: Optimal caching time $\widetilde{T_C^*}$ under different (α, β) configurations. In proactive caching analysis, the overestimation of outbreak of the epidemic content advances the optimal caching time in general, comparing to Fig. 4. $N = 1000$, $I_0 = 1$, $L = 100$, $\delta = 1$, $\nu = 1$, $\kappa_1 = 0.1$, $\kappa_2 = 0.2$, $\kappa = 0.15$, $\tau = 1$, $T_i = 0$, $T_f = 150$, $\Lambda_I(0) = 20$, $\Lambda_R(0) = 10$.

the cost of caching is the major determinant of optimal caching time in proactive caching analysis, for it always overestimates the severeness of the outbreak.

Finally, we plot the optimal caching time T_C^* versus virality ν in Fig. 6, for virality is proposed as another content feature rather than popularities in this paper. Obviously, a highly viral content advances the optimal caching time to handle suddenly massive requirements; while for lowly viral content, the optimal caching time is late since it takes longer time for the content to infect enough users to help serve the users. For different situations of caching cost and the requirement of system dynamic, our previous discussion still holds; that is, large α delays optimal caching time and large β advances it. The virality of an epidemic content has more pragmatic meaning than merely the popularity in a mobile social network [12], since prefect and centralized traffic monitoring to obtain popularity is often arduous and not timely. Therefore, virality holds a chance as a more realistic feature when it comes to dynamic social systems, and an important system characteristic when designing communication and information sharing networks. It is obvious that for a small change in virality, *e.g.* from 1.5 to 2.0, the optimal caching time could varies from about 36 to 30 seconds. Consequently, virality prediction [12] becomes a critical information collection mechanism of content properties in mobile social networks for caching utilization.

VI. Conclusion

In this paper, we investigate the optimal utilization of caching in mobile social networks for content dissemination by deriving optimal caching time, compromising between the system cost of caching and the system service requirements, which are two important factors when designing networks for efficient content dissemination. Via epidemic modelling, we exploit optimal control theory to determine the optimal caching time for efficacious and timely control. Subsequently, as a preventive system response to predict and handle a severe outbreak of demand for the content, we provide proactive caching analysis to alleviate such instantaneous service burden for the infrastructure in the network. Analytical derivation

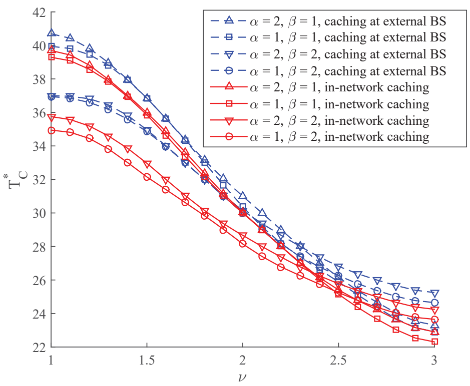

Fig. 6: Optimal caching time versus virality ν under different (α, β) configurations. A highly viral content leads to early optimal caching time to handle massive outbreak. Virality hence provides an important content feature when implementing caching. $N = 1000$, $I_0 = 1$, $L = 100$, $\delta = 1$, $\kappa_1 = 0.1$, $\kappa_2 = 0.2$, $\kappa = 0.2$, $\tau = 1$, $T_i = 0$, $T_f = 150$, $\Lambda_I(0) = 20$, $\Lambda_R(0) = 10$.

and simulation results support that the epidemic content can be efficiently spread through service provided by caching at external BS, while cooperative in-network caching boosts the dissemination of the epidemic content by storing the content in users and sharing it upon contacts with users still waiting for the content. Moreover, virality is shown to be an important content feature when implementing caching in dynamic systems. Consequently, from viewpoints of system dynamic and costs, this paper reveals novel insights and analysis of optimal caching time toward effectual and robust epidemic content dissemination in mobile social networks.

REFERENCES

[1] A. M. Kaplan, "If you love something, let it go mobile: Mobile marketing and mobile social media 4x4," *Business horizons*, vol. 55, no. 2, pp. 129–139, 2012.

[2] L. Wang, H. Wu, W. Wang, and K.-C. Chen, "Socially enabled wireless networks: resource allocation via bipartite graph matching," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 128–135, 2015.

[3] D. J. Daley, J. Gani, and J. M. Gani, *Epidemic modelling: an introduction*. Cambridge University Press, 2001, vol. 15.

[4] H. Sun and C. Wu, "Epidemic forwarding in mobile social networks," in *Proc. of IEEE ICC*, 2012.

[5] Q. Xu, Z. Su, K. Zhang, P. Ren, and X. Shen, "Epidemic information dissemination in mobile social networks with opportunistic links," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 3, pp. 399–409, 2015.

[6] P.-Y. Chen, S.-M. Cheng, and K.-C. Chen, "Optimal control of epidemic information dissemination over networks," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2316–2328, 2014.

[7] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *arXiv preprint arXiv:1408.2701*, 2014.

[8] Y. Wang, J. Wu, and M. Xiao, "Hierarchical cooperative caching in mobile opportunistic social networks," in *Proc. of IEEE GLOBECOM*, 2014.

[9] X. Zhuo, Q. Li, G. Cao, Y. Dai, B. Szymanski, and T. L. Porta, "Social-based cooperative caching in dtns: A contact duration aware approach," in *Proc. of International Conference on Mobile Adhoc and Sensor Systems (IEEE MASS)*, 2011.

[10] L. Han, M. Punceva, B. Nath, S. Muthukrishnan, and L. Iftode, "Socialcdn: Caching techniques for distributed social networks," in *Proc. of International Conference on Peer-to-Peer Computing (IEEE P2P)*, 2012.

[11] Ericsson. (2014) Ericsson mobility report. [Online]. Available: http://www.ericsson.com/mobility-report

[12] L. Weng, F. Menczer, and Y.-Y. Ahn, "Virality prediction and community structure in social networks," *Scientific reports*, vol. 3, 2013.

[13] H. Hsu and K.-C. Chen, "A resource allocation perspective on caching to achieve low latency," *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 145–148, 2016.

[14] D. E. Kirk, *Optimal control theory: an introduction*. Courier Corporation, 2012.

[15] S. Meyn, *Control techniques for complex networks*. Cambridge University Press, 2008.

[16] L. S. Pontryagin, *Mathematical theory of optimal processes*. CRC Press, 1987.

[17] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the levy-walk nature of human mobility," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 630–643, 2011.

[18] S. Kim, C.-H. Lee, and D. Y. Eun, "Superdiffusive behavior of mobile nodes and its impact on routing protocol performance," *IEEE Trans. Mobile Comput.*, vol. 9, no. 2, pp. 288–304, 2010.