C4: A New Paradigm for Providing Incentives in Multi-hop Wireless Networks

Chi Zhang*, Xiaoyan Zhu†, Yang Song*, and Yuguang Fang*†

*University of Florida, †Xidian University

Abstract— **For a multi-hop wireless network (MWN) consisting of mobile nodes controlled by independent self-interested users, incentive mechanism is essential for motivating mobile nodes to cooperate and forward packets for each other. Existing solutions such as barter based, virtual-currency based and reputation based schemes are either less effective or incur high implementation costs, and therefore do not fit well with the unique requirements of MWNs. In this paper, we propose a novel and promising incentive paradigm, *Controlled Coded packets as virtual Commodity Currency* (C4), to induce cooperative behaviors in MWNs. In our C4, through introducing several techniques from network coding, coded information packets are utilized as a new kind of virtual currency to facilitate packet/service exchanges among self-interested nodes in a MWN. Since the virtual currency implemented in this way also carries useful data information, it is the counterpart of the so-called commodity currency in the physical world, and the overhead brought by C4 is extremely small compared to traditional schemes. We theoretically show that C4 is perfectly efficient to support MWNs with broadcast and multicast traffics. For pure unicast communications, by adjusting the grouping parameter, our C4 provides a systematic way to smoothly trade incentive effectiveness for implementation cost, and traditional barter based and virtual-currency based schemes are just two extreme cases of C4. We also show that when our C4 is combined with the social network formed by mobile users in the MWN, the implementation costs can be further reduced without sacrificing incentive effectiveness.**

I. INTRODUCTION AND MOTIVATION

In a multi-hop wireless network (MWN), when the source and the destination nodes are not within direct transmission range of each other, they must rely on intermediate nodes to forward packets for them. Hence, the performance of a MWN heavily depends on the participating nodes' willingness to *cooperate*. If all nodes are cooperative, such as in military networks configured by a central authority, cooperation can be taken for granted. However, for most current and emerging MWNs, participating nodes are owned and administered by different authorities such as different persons, and therefore are *autonomous*. When a node forwarding traffic for other nodes, it expends its own bandwidth and power resource without any direct benefit. A *self-interested* node therefore has a strong incentive to *free ride*, i.e., use the network resources of other nodes without contributing its own. If free-riding behaviors prevail, such networks even cannot function. Therefore, the proper design of *incentive mechanism* for encouraging resource sharing at the network layer is essential for the success of any MWN in civilian or commercial environments [1].

Obviously, the interactions among autonomous and self-interested entities can be modeled and analyzed as a socio-economic system, and how to stimulate cooperative behaviors in such a system is an extensively studied topic in sociology and economics with a rich collection of analyzing techniques and promising solutions [2], [3]. Therefore, it is not surprising that all proposed incentive mechanisms for MWNs in the literature draw analogies from their counterparts in human

TABLE I
THE DESIGN SPACE OF INCENTIVE MECHANISMS

Economic mechanisms	Bartering	Indirect bartering	Commodity currency	Fiat currency
Mechanisms used in wireless networks	Bartering (tit-for-tat) [4]–[8]	Networked bartering [9]	?	Virtual currency [10]–[14]
		Generous tit-for-tat [15]	Reputation [16]–[18]	
Social mechanisms	Direct reciprocity	Indirect reciprocity	Reputation	

societies. Table I gives a glimpse of design space of incentive mechanisms for MWNs and points out their relationships with economic and social mechanisms enabling cooperations in human societies. Existing approaches for providing incentives basically can be classified into three categories as follows.

The first category is *barter* based approaches [4]–[8], which are based on *direct reciprocity*: node A would provide resources/services for node B only if B simultaneously provides resources/services for node A. This kind of *bilateral* and *synchronous* resource/service trading makes barter extremely simple to implement. From a system perspective, there is no need to keep any long-term state information, in the form of either reputation or currency, and as a consequence the implementation cost of barter is almost zero. However, synchronous trading is easy to fail when an action and its reward are not simultaneous [9], [15]. The second category is *virtual-currency* based [10]–[14], in which participating nodes would earn virtual currency by providing resources/services to others and spend the virtual currency to obtain resources/services from others. By taking virtual currency as a medium of exchange, nodes can then trade resources/services asynchronously, which overcomes the shortcoming of barters. Virtual currency, however, incurs a high implementation overhead, e.g., billing and e-cash transfers, implementations of centralized bank and dispute-resolution mechanisms, etc. In the third category, i.e., *reputation* based approaches [16]–[18], participants build up their reputation scores by providing services for others, and highly reputed participants receive preferential treatment when they need help. Obviously, reputation scores here can be treated as another form of virtual currency. Therefore, reputation based approaches share the same advantages and disadvantages as virtual-currency based ones. To sum up, existing incentive mechanisms are either less effective or incur high implementation costs, and therefore do not fit well with the requirements of MWNs. A new design paradigm is needed.

Before introducing any new solution, the first question we should ask is: *Is there any room for further exploration*? If we go through the design space of incentive mechanisms, there do exist an unexplored area which is the counterpart of commodity currency in the physical world (i.e., the question mark in Table I). Economists define *currency* (or *money*) as a generally acceptable medium of exchange [2], [3]. With the use of currency, the problem of "double coincidence of wants" is avoided, and the *transaction cost* of searching a successful trade is reduced. Historically, currency originated as *commodity currency*. When a physical commodity (e.g., compressed tea leaves in ancient China) has value to everyone and is

TABLE II
COST ANALYSIS OF DIFFERENT INCENTIVE MECHANISMS

Scheme / Cost	In the physical space			In the digital space		
	Barter	CC	FC/Rpt	Barter	C4	VC/Rpt
Transaction	High	Low	Low	High	Low	Low
Transportation	Zero	High	Low	Zero	Low	Low
Implementation	Zero	Zero	Low	Zero	Low	High

(a) Multi-hop wireless network   (b) Network model for performance analysis

Fig. 1. **A generic architecture for multi-hop wireless networks.**

generally accepted as the payment for other goods/services, it can serve as a currency. This kind of currency has *intrinsic value* (i.e., valuable in its own right), and still can be consumed as a physical commodity (i.e., to make tea) when not needed for trade. The problem of commodity currency like tea leaf bricks is that it is heavy and hard to transport from one place to another. *Fiat currency* (paper currency) is invented to reduce the costs involved in storing and carrying commodity currency in trading (i.e., *transportation cost*). Fiat money is without intrinsic value as a physical commodity, and derives its value by being declared by a government to be legal tender. Obviously, virtual currency discussed above (e.g., e-cash) is exactly the counterpart of fiat currency in the physical word.

So, our second question comes: *What is the counterpart of commodity currency in the digital world*? There are only two conditions for a likely candidate: (i) it should carry data information because the commodity in a MWN can only be data packet; (ii) it has value to everyone. Following these clues, the key idea of our new incentive paradigm, *Controlled Coded packets as virtual Commodity Currency* (**C4**), naturally emerges: *coded packets can serve as virtual commodity currency for MWNs*. By utilizing network coding, original data packets for different mobile users are mixed to produce coded packets. As a consequence, each coded packet has an intrinsic value to every node involved, and is ready to act as virtual currency to facilitate cooperations.

The only question left then is: *What are the benefits of using virtual commodity currency (i.e., C4) in MWNs?* To fairly evaluate an incentive mechanism, we need consider three kinds of costs. As described in Table II, in the physical space, compared to the exchange of physical commodities, the *implementation cost* of fiat currency (FC) or the mouth-to-mouth reputation (Rep) can be ignored. Barter has a high transaction cost; commodity currency (CC) has a high transportation cost; only FC can keep all three costs low, and therefore is the best choice. This explains the fact that nearly all contemporary economic systems are based on fiat currency. What happens in the digital space like a WMN? The key difference is that the goods in a WMN are information packets. Virtual currency (VC) and reputation are also stored and transacted in the form of information bits. Therefore, compared to the data packet exchanges, the implementation cost of VC/Rep is pretty high. The implementation cost of **C4** is always smaller than that of VC, because when e-cash is used as the medium of exchange, it only represents control overhead; while when coded packet is used, it also carries useful data. Therefore, from Table II, we can see that **C4** is the best choice in the digital space.

The rest of the paper is organized as follows. The design of **C4** for a generic MWN is described in Section II. In Section III we evaluate the performance of **C4** through theoretical analysis and simulations. In Section IV, we show how to utilize the social network formed by mobile nodes to further improve the performance of **C4** with pure unicast communications. Conclusion and future work are described in Section V.

II. DESIGN AND IMPLEMENTATION OF OUR C4

In this section, we describe the design of our **C4**. We first clarify the target scenarios and basic assumptions.
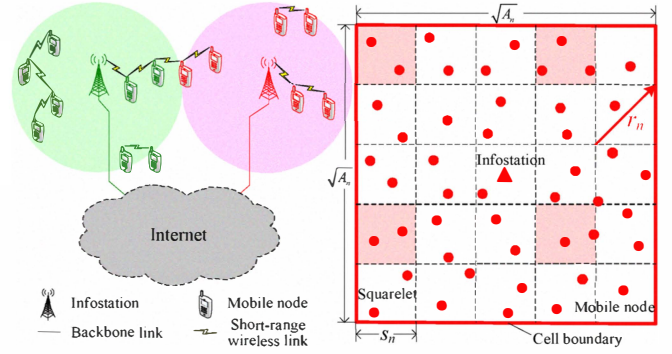
A. System Model and Problem Formulation

Based on the common requirements of future mobile communication environment [19]–[22], our **C4** assumes the following generic model for MWNs. As illustrated in Fig. 1(a), there are two kinds of entities: mobile nodes and infostations.

• *Mobile nodes* are controlled by autonomous and self-interested clients and are interested in Internet access services. A mobile node can establish a *short-range wireless link* (e.g., Wi-Fi) with other mobile nodes in its vicinity. The short range links tend to be intermittent because of node mobility.

• *Infostations* are managed by the system operator and directly connected to the Internet with reliable and high-bandwidth links (i.e., *backbone links*). An infostation can use a long-range low-bandwidth radio (e.g., cellular interface) to connect with a remote mobile node (but we do not assume that all mobile nodes are covered by infostations), or use a short-range wireless link with high data rate to connect with a close mobile node. Infostations are the data sources within the wireless domain and want to provide better services to clients.

As illustrated in Table III, this generic model includes several important multi-hop wireless network architectures which attract great interest from both academia and industry.

We assume that all mobile nodes are self-interested and still *rational*. They have the non-cooperative behaviors mainly because they want to save resource such as bandwidth and battery power. We also assume that all infostations are under the control of one authority, and they will do all they can to encourage mobile nodes to use short-range links.

Unlike previous work [10]–[14], [16]–[18], [23], [24], our **C4** does not make any assumptions about routing protocols used in MWNs. Our **C4** is designed to support traditional store-and-forward routing schemes as well as new routing schemes in a DTN fashion (i.e., store-carry-and-forward [25]). Also our **C4** allows all possible combinations of traffic patterns (e.g., broadcasts, multicasts and unicasts). The only requirement is that most traffics go through infostations.

B. Methodology of Our C4

We take one broadcast session from the infostation to mobile nodes as an example to illustrate the basic design of our **C4**.

Suppose this broadcast session is about distributing a movie stored in a web server in the Internet. The infostation treats all network layer packets received from the web server as *original data packets* (OPs). We assume each OP has l bytes. At the infostation, *random linear network coding* (RLNC) [26] is applied to a finite set of k OPs (i.e., OP_1, OP_2, ..., OP_k), which is called a *generation*. Then each generation can be regarded as a $k \times l$ matrix OP, with rows being the k OPs of the generation, and columns the l bytes of each OP. The encoding operation produces a linear combination of the OPs by $CB = CV \cdot OP$, where CV is an $k \times k$ matrix composed of randomly selected

TABLE III
EXAMPLES OF MULTI-HOP WIRELESS NETWORKS

Multi-hop wireless network	Infostation	Mobile node
Multi-hop cellular network [19]	Base station	Mobile station
Wireless mesh network [20]	Mesh router	Wireless client
Mobile social network [21]	Service provider	Mobile user
Vehicular ad hoc network [22]	Roadside info. unit	Vehicle

Fig. 2. **A comparison between bartering (without coding) and our C4.**

(a) Bartering (b) Random linear network coding

coding coefficients in the Galois field $\mathbb{GF}_q$ of size q. The *coded data blocks* (rows in CB) and the *coding vectors* (rows in CV) are concatenated as the *coded data packets* (CPs). For example, the coded data block $\mathtt{CB}_j = \sum_{i=1}^{k} \mathtt{CV}_{ji} \cdot \mathtt{OP}_i$ where $\mathtt{CV}_{ji} \in \mathbb{GF}_q$ and the coded packet $\mathtt{CP}_j = \mathtt{CV}_j \,\|\, \mathtt{CB}_j$. Two coded packets $\mathtt{CP}_i$ and $\mathtt{CP}_j$ are called *independent* if $\mathtt{CV}_i$ and $\mathtt{CV}_j$ are independent vectors. In our **C4**, instead of sending OPs, the infostation sends CPs to mobile nodes with short-range links. The decoding operation at the destination (i.e., mobile nodes), in its simplest form, is the matrix multiplication $\mathtt{OP} = \mathtt{CV}^{-1} \cdot \mathtt{CB}$, where each row of CB represents a coded data block and each row of CV represents the coding vectors accomplished with it. The successful recovery of the original packets OP requires that the matrix CV be of full rank, i.e., the destination must collect k independent CPs.

Each mobile node can either download packets from an infostation or exchange packets with neighboring mobile nodes. In a non-cooperative network without any incentive mechanisms, the former is the only mechanism for packet dissemination. It only uses the high-speed channel between an infostation and a node near it, while wasting all the equally excellent channels between neighboring mobile nodes. A barter-like scheme (without network coding) alleviates this problem as follows: when two mobile nodes contact, they inspect the packet contents of each other. If each node identifies a packet that it wants, a bilateral packet exchange takes place. However, even for broadcasts, nodes can easily end up in a deadlock situation because of the requirement of mutual wants. In our **C4**, let $S_u(t)$ and $S_v(t)$ denote the subspaces spanned by the CPs at neighboring node u and v, respectively, at the beginning of the timeslot t. If $S_u(t) \not\subseteq S_v(t)$, we say node v wants CPs at u. If there exists mutual wants, then node u and v can successfully exchange CPs as follows: node u sends a new CP $p_u \in S_u(t)$ and $p_u \notin S_v(t)$, and node v sends back a new CP $p_v \in S_v(t)$ and $p_v \notin S_u(t)$ as return.

The design of **C4** relies on two unique features of RLNC:

• *RLNC can greatly improve the possibility of successful exchange.* As proved in Lemma 2.1 in [27], when RLNC is used, the probability that two randomly selected nodes have the mutual wants tends to 1 when q, the size of $\mathbb{GF}_q$, is large enough (e.g., $q = 2^8$). Fig. 2 illustrates this effect. Suppose there are only two mobile nodes, and three OPs need be broadcasted at the infostation. It is easy for two nodes run into a deadlock situation as exemplified in Fig. 2(a). When virtual currency is introduced, node B can buy packet 2 from node A by sending back an e-cash at timeslot 3. However only one data packet is transmitted, and e-cash transfer is overhead. By mixing the OPs, because with high probability each CP brings some new information, it can be treated as virtual commodity currency, i.e., generally acceptable in payment. For example, in Fig. 2(b), at timeslot 3, node B can buy a packet from A by paying a CP $(0, 1, 1) \cdot \mathtt{X}$. After timeslot 3, node A can decode all OPs, and node B only need one more CP. High exchange efficiency achieves without incurring any overhead.

• *RLNC can serve as a free cipher.* Before the destination obtains enough independent CPs, it cannot recover all OPs, even thou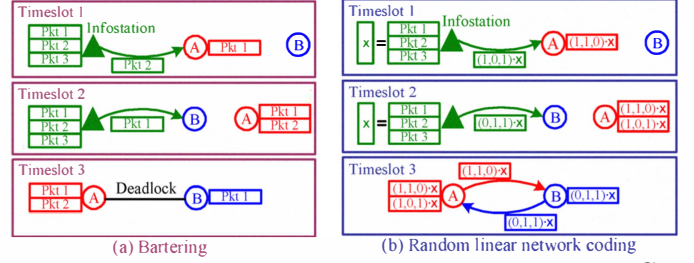gh every CP contains some information about every OP. Therefore, RLNC can serve as a free cipher for a time period. This functionality is extremely useful for supporting unicasts. Also take Fig. 2(b) as an example. Now suppose three OPs are destined for different mobile nodes (A or B). Before node A receives three independent CPs, he cannot tell which packet is destined to whom. Therefore, even when three OPs are all destined to node B, node A still has incentive to participate in exchanges before he collects three independent CPs. In this way, we enforce node A to help relay one CP to node B at timeslot 3, which is impossible for bartering.

C. Implementation Details of Our C4

According to source-destination relationship within the wireless domain, all original data packets in our **C4** can be classified into two categories: (1) *download data packets*: the packets from the infostation to mobile nodes; and (2) *upload data packets*: the packets from mobile nodes to the infostation.

Our **C4** take the whole network layer packet as an original data packet (including packet header). Each original data packet has a unique packet ID, and a session ID indicating which session this packet belongs to. A mobile node can tell whether he/she is interested in it by checking its session ID. In our **C4**, every transmitted packet within the wireless domain is a coded packet, and has a special format. At the infostation, some download data packets are selected and coded to produce virtual commodity currency (VCC). We call these coded data packets as *VCC packets*. As illustrated in Fig. 3(a), a coded packet consists of a header and a body. The body stores the coded data block. For a VCC packet, the header consists of three parts: (1) TPE field, which is set to 11 to indicate a VCC packet; (2) session field, which indicates session IDs involved in the coded data block in the body part, and it can be implemented by a Bloom filter; (3) coding-vector field, where k is the generation size, p is the number of OPs encoded in the packet and c_i is the coding coefficient related to the OP with packet ID oid_i for $1 \le i \le p$. Unlike traditional RLNC packet, here $k \ne p$. The reason will be explained a little later. A VCC packet $\mathtt{CP}_{VCC}$ is valid for a mobile node u, only if u is interested in at least one session involved in $\mathtt{CP}_{VCC}$ and does no collect enough coded packets to recover all OPs of this session. Download/upload data packets will not be directly transmitted in our **C4**. They will be carried by VCC packets. For a download data packet with packet ID oid_q and destination node ID d_id, which is not selected to produce VCC packet, the infostation first selects one $\mathtt{CP}_{VCC}$ as the carrier, and then combines these two packets as shown in Fig. 3(b). The TPE field is set to 01 to indicate a *coded download packet*. The coded block is the linear combination of the coded block of $\mathtt{CP}_{VCC}$ with the coding coefficient 1 and the download data packet with the coding coefficient c_q. At a mobile node with node ID s_id, the *coded upload packet* is constructed in a similar way, as illustrated in Fig. 3(c). The TPE field is set to 00 to indicate a coded upload packet and the source node ID s_id is also included in the packet header.

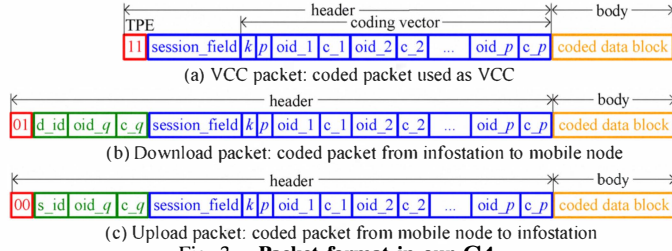There are only two kinds of elementary interactions when

(a) VCC packet: coded packet used as VCC

(b) Download packet: coded packet from infostation to mobile node

(c) Upload packet: coded packet from mobile node to infostation

Fig. 3. **Packet format in our C4.**

two mobile nodes u and v contact in our **C4**: (i) *exchanges of coded packets*: When node u has a coded packet which v is interested in, and the same thing happened for node v, then u and v have incentives to exchange these packets. (ii) *exchanges of relay service and VCC packets*: When node u wants node v to relay an upload/download data packet, and node u has at least two valid VCC packets for v, then in this interaction, node u sends two packets to v: the coded upload/download packet needs to be relayed and one valid VCC packet as the payment. The upload/download data packet here should be carried by another valid VCC packet for v. Note that if the upload/download data packet is directly sent to v, then node v can drop this packet a little later without being detected. By combining the data packet with the VCC packet, our **C4** provides incentives for relay node v to keep this data packet before v can decouple it from the VCC packet. By decomposing all possible end-to-end communications into a series of elementary interactions, our **C4** puts no constraints on the underlying routing protocols or traffic patterns, and therefore achieves the maximum flexibility in supporting different application scenarios in MWNs.

The coding strategy adopted by infostations to generate VCC packets determines the performance of our **C4**. The key point is how to select sessions that need be mixed in generating VCC packets. There exists a tradeoff: if we mix the sessions interested by different mobile nodes, we can obtain VCC packets which are valid for a larger population of mobile nodes. Then the VCC packets can achieve a better exchange efficiency. However, the useful data information contained in the VCC packet for a mobile user is also decreased, which means the overhead of using VCC packets becomes larger. Therefore, a cautious choice should be made in order to obtain an optimized balance. A quantitative analysis of optimized coding strategies will be presented in Section III in detail. Here, we just give two qualitative guidelines. (i) Because all users are interested in broadcast data packets, they can be safely mixed without worrying about overhead. The only problem is that in order to prolong the valid time of VCC packets, the size of one generation, i.e. k, should be large enough. For traditional RLNC, this means the size of the packet header is pretty large, because the size of coding vector which should be included in the coded packet header is $\Theta(k)$. To avoid this problem, our **C4** utilizes sparse RLNC proposed in [28] to control overhead. Given k original data packets as one generation, sparse RLNC first randomly chooses p data packets, then performs RLNC on these p packets to generate a coded packet. It can be shown that the successful exchange probability is on the same order of that of RLNC, while the size of the packet header is $\Theta(p)$ where $p \ll k$. Note that our packet format shown in Fig. 3 already includes the unique feature of sparse RLNC by embedding k, p and OP IDs into coding-vector field. (ii) When only unicast traffics exist in a MWN, we have to mix OPs for different users to generate valid VCC packets. When k OPs for different sessions are grouped for one generation and coded with sparse

RLNC, the session field in a VCC packet is determined by k OPs in the generation, not by p OPs selected for this VCC packet. However, it is possible that a VCC packet does not contain any useful data information for a mobile user, but that user still treats it as valid. A mobile node even cannot find this before he/she can recover all original packets. Therefore, the infostation can intentionally do this to provide incentives or optimize the whole system's performance. Our **C4** only requires the long-term fairness among mobile nodes and utilizes the free cipher functionality provided by RLNC to enforce this kind of operations.

III. PERFORMANCE ANALYSIS OF OUR C4

A. Network Model for Performance Analysis

In order to make the quantitative study possible, we detail our network model proposed in Section II-A as follows.

[Cell Model] Consider a square geography of area A_n with a fixed infostation at the center, as shown in Fig. 1(b). We assume the geography wraps around each boundary, effectively creating a torus. We refer to this torus as a *cell*. A cell is intended to mimic a typical multi-infostation network in which an infinite grid of infostations populate an infinite plane. The area A_n relative to the single infostation serves to characterize the density of fixed infostations over the terrain.

[Mobility Model] The cell is populated with n mobile nodes with independent mobility processes as follows. We first divide the cell into $m = \Theta(n)$ squarelets of area s_n^2 each, resulting in a two-dimensional $\sqrt{m} \times \sqrt{m}$ discrete torus (we assume $\sqrt{m}$ is an integer). We assume the time is slotted and each node independently performs a simple random walk on the two-dimensional $\sqrt{m} \times \sqrt{m}$ discrete torus, i.e., let a node be in squarelet s at timeslot t, then, at timeslot $t+1$, the node is equally likely to be in the same squarelet s or any of the four adjacent squarelets (i.e., up, down, left, and right squarelets).

[Protocol Model] For characterizing the condition for a successful wireless transmission, we adopt the *protocol model* proposed in [29]. We assume that all mobile nodes (including the infostation) use a common range r_n for their short-range transmissions, and a transmission from node i to node j is successful if and only if $d_{ij} \leq r_n$ and $d_{kj} \geq (1+\Delta)r_n$ for any other simultaneous transmitter, say node k. Here, d_{ij} is the distance between nodes i and j, and Δ is a positive constant independent of n. We assume that when two nodes are in the same squarelet, they are *neighbors*, i.e., they can establish a short-range wireless link. Therefore, $r_n = \sqrt{2}s_n$.

[Transmission Scheduling] A squarelet is called *active* at timeslot t, if the MAC scheduling scheme allows one node in that squarelet to transmit at timeslot t. Based on protocol model, we can guarantee that there exists an interference-free schedule such that each squarelet becomes active regularly once in L timeslots and it does not interfere with any other simultaneous active squarelets. Here L depends only on Δ, and is independent of n (cf. Proposition 1 in our previous work [30]). The shaded squarelets in Fig. 1(b) illustrate an example of a group of simultaneous active squarelets when $L = 9$. For an active squarelet, at most one node pair will be scheduled to communicate. When the squarelet with the infostation (i.e., *central squarelet*) is active, the infostation will be one party of that node pair and another party will be randomly chosen from all nodes in that squarelet. For any other squarelet (i.e., *regular squarelet*), when it is active, a node pair is randomly selected from all possible node pairs in that squarelet. This is intended to mimic the behaviors of IEEE 802.11-like MAC protocol. For each timeslot, we assume
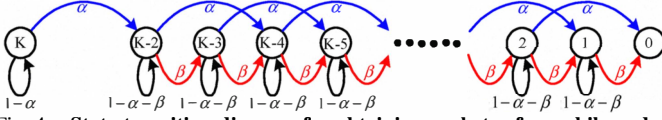
Fig. 4. **State transition diagram for obtaining packets of a mobile node.**

only two packets can be transmitted, i.e., just enough for one successful elementary interaction.

We note that above model is the standard network model widely used in the literature for MWN performance analysis (see [29]–[31] and references therein), and its behaviors are characterized by the following Lemma when $n \to \infty$.

Lemma 1: (1) Given $r_n = \sqrt{2A_n/m} = \Theta(\sqrt{A_n/n})$, the network formed by short-range wireless links is disconnected, i.e., there does not exist a contemporaneous path between two random selected nodes with high probability. (2) For a given timeslot, the probability that a mobile node is in the active central squarelet and selected into the communication pair is $p_I = \frac{\gamma}{m \cdot L} = \Theta(1/n)$, where γ is a constant. (3) For a given timeslot, the probability that a mobile node is in an active regular squarelet and selected into the communication pair is μ, and μ is a constant, i.e., independent of n.

B. Performance Analysis for Broadcast and Multicast Traffics

We first consider broadcast scenarios. Suppose the infostation has K data packets to be distributed to n mobile nodes. For this scenario, in the central squarelet the elementary interaction is for the infostation to send two packets to a mobile node, while in regular squarelets the elementary interaction is packet exchanges between two mobile nodes.

The *effectiveness* of incentive mechanism IM is measured by the *expected packet delivery time*, i.e., $T_D(IM)$. For broadcast scenarios, let T_{IM} denote the time for all mobile nodes to obtain K data packets under the mechanism IM, then $T_D(IM) = \frac{\mathbb{E}[T_{IM}]}{K}$. The smaller the time $T_D(IM)$, the more effective the mechanism IM. The time $T_D(IM)$ for different mechanism IM is characterized by the following Theorem:

Theorem 1: [$T_D(IM)$ **for Broadcasting K Data Packets**] When all nodes are fully cooperative and no network coding scheme is used (i.e., $IM = C$), then $T_D(C) = \Theta\left(\log n + \frac{n}{K}\right)$. When all nodes are non-cooperative, and (1) if bartering is used (i.e., $IM = B$), then $T_D(B) = \Theta\left(\log^2 n + \frac{n}{K}\right)$; (2) if virtual currency is used (i.e., $IM = VC$), then $T_D(VC) = \Theta\left(\log n + \frac{n}{K}\right)$; (3) if our **C4** is used with $k = K$ and p is a constant (i.e., $IM = C4$), then $T_D(C4) = \Theta\left(1 + \frac{n}{K}\right)$.

Proof: We observe that for a given timeslot, whether a mobile node can obtain a new packet depends on two conditions: (i) Whether this node is selected into a communication pair? This condition is characterized by Lemma 1 and is independent of the incentive mechanism used. (ii) When this node is selected into a communication pair, whether it can make a successful exchange with its partner? Here, we assume that for an arbitrary node pair, there exists a probability p_E of successful exchange which only depends on the incentive mechanism. We want a static, summarized characterization of p_E, i.e., we obtain a value of p_E which is averaged over all possible node pairs and all timeslots. We then simplify our analysis by assuming that given the incentive mechanism, the probability of successful exchange for any node pair at any timeslot is the same as this value. It is not hard to see that this key assumption is inconsistent with the interaction dynamics. Nevertheless, our simulation results agree closely with the analytical results, indicating that this assumption works well in systems with moderately large number of nodes $n \geq 50$.

Based on this assumption, we can model the dynamics of obtaining packets for a given mobile node as a discrete time

Markov chain illustrated in Fig. 4. Denote the state as the number of packets remaining to be obtained by a mobile node. Initially a node is at state K. Since the first two packets must be obtained from an infostation, the next state is $K-2$. Subsequently, in states $s \in \{1, \cdots, K-2\}$, there are three possibilities in each timeslot:

- With probability α, the node is in the active central squarelet and obtains two packets from the infostation. The state goes from s to $s-2$.
- With probability β, the node is in an active regular squarelet and obtains one packet from another mobile node. The state goes from s to $s-1$.
- With probability $1-\alpha-\beta$ no new packets are obtained because the node is not selected into communication pairs or because the node cannot make a successful exchange with its partner and the state stays the same.

For α, because the infostation always has the incentive to send packets to mobile nodes, this probability is independent of the incentive mechanism and from Lemma 1(2) we know that $\alpha = p_I = \Theta(1/n)$. For β, obviously $\beta = \mu \cdot p_E = \Theta(p_E)$.

We denote the expected first passage time from state i to state 0 as PT_i, where $2 \leq i \leq K-2$. Conditioning on the next state transition and rearranging yields the difference equation,

$$PT_i = \frac{1}{\alpha + \beta} + \frac{\beta}{\alpha + \beta}PT_{i-1} + \frac{\alpha}{\alpha + \beta}PT_{i-2}$$

where the boundary conditions are given by $PT_0 = 0$ and $PT_1 = \frac{1}{\alpha+\beta}$. Using z-transforms, we can obtain

$$PT_i = \frac{i(\beta + 2\alpha) + \left(1 - \left(\frac{-\alpha}{\alpha+\beta}\right)^i\right)\alpha}{(\beta + 2\alpha)^2}$$

It is obvious that $\mathbb{E}[T_{IM}] = 1/\alpha + PT_{K-2}$, where $1/\alpha$ is the expected time until a node first obtains two packets from the infostation. When $n \to \infty$, we can obtain

$$\mathbb{E}[T_{IM}] \to \frac{K}{\beta} + \frac{1}{p_I} = \frac{K}{\mu \cdot p_E} + \frac{1}{p_I}.$$

When all mobile nodes are fully cooperative and no network coding techniques is used, p_E is the probability that a node has a new packet for another node. Using the result about coupon collecting problem [27], we have $p_E = \Theta(1/\log n)$. Therefore, $T_D(C) = \Theta\left(\log n + \frac{n}{K}\right)$.

When mobile nodes are non-cooperative, we have three possibilities: (1) For bartering, p_E is the probability that two nodes can provide a new packet for each other, thus $p_E = \Theta\left(1/\log^2 n\right)$ and $T_D(B) = \Theta\left(\log^2 n + \frac{n}{K}\right)$. (2) For VC, we assume each node has enough VC. Therefore $T_D(VC)$ is the same as $T_D(C)$ in cooperative case. (3) For our **C4**, using property of RLNC [26], we have $p_E \geq (1 - 1/q)^2 = \Theta(1)$. Therefore, $T_D(C4) = \Theta\left(1 + \frac{n}{K}\right)$. ∎

The *cost efficiency* of incentive mechanism IM is measured by the *expected number of control packets* (i.e., $C_P(IM)$) that need be transmitted for obtaining one data packet. For broadcast scenarios, let C_{IM} denote the expected total number of control packets transmitted for one mobile node to obtain K data packets under the mechanism IM, then $C_P(IM) = \frac{\mathbb{E}[C_{IM}]}{K \cdot n}$. The smaller the number $C_P(IM)$, the more cost-efficient the mechanism IM. The number $C_P(IM)$ for different IM is characterized by the following Theorem:

Theorem 2: [$C_P(IM)$ **for Broadcasting K Data Packets**] (1) $C_P(C) = C_P(B) = C_P(C4) = 0$ and (2) $C_P(VC) = \Theta\left(1 - \frac{\log n}{n}\right)$. (The proof is omitted for the reason of space.)
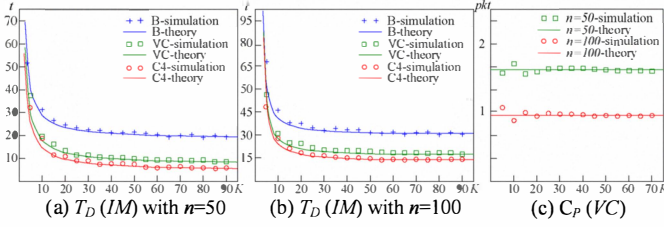
Fig. 5. $T_D(IM)$ **and** $C_P(IM)$ **as functions of** K.

In Fig. 5, we compare theoretical results obtained from Theorem 1 and 2 with simulation results when $L=4$. We show $T_D(IM)$ and $C_P(IM)$ averaged over 100 simulation runs for different IMs. The number of nodes is held constant at $n=50$ ($m=25$) for Fig. 5(a) and at $n=100$ ($m=64$) for Fig. 5(b), while the number of data packets K is varied. Fig. 5(c) shows $C_P(VC)$ for $n=50, 100$. The results about $C_P(C)$, $C_P(B)$, and $C_P(\mathbf{C4})$ are all zeros, and are not shown in the figure. In all cases, the differences between the simulation results and our theoretical results are found to be very small.

From Theorem 1 and 2 and simulation results, we can make the following observations about broadcast scenarios:

- For all possible values of K, the packet delivery time required by our **C4** is always smaller than that required by other possible incentive mechanisms. Our **C4** is even better than the case when all nodes are cooperative but no coding is used. All these improvements are obtained with zero cost. Therefore, our **C4** is the best choice for broadcast traffics.

- We assume the number of data packets which need be broadcasted is pretty large, i.e., $K = \Omega(n)$. In this regime, from Theorem 1, we have $T_D(B) = \Theta\left(\log^2 n\right)$, $T_D(VC) = \Theta(\log n)$, and $T_D(\mathbf{C4}) = \Theta(1)$. The improvements converge to a factor on the order of $\log n$ which is independent of K.

- It is believed in previous work [4]–[8] that for broadcast traffic, bartering is better than virtual-currency based schemes. Our results also provide an analytical evidence to support this belief because compared to bartering, the VC based scheme only provides an $\Theta(\log n)$ improvement to the packet delivery time, with a cost of $\Theta(n)$ control packets per node.

We then consider multicast scenarios. The following corollary shows that the same performance can be achieved.

Corollary 1: Suppose the infostation has K data packets to send to $\epsilon \cdot n$ randomly selected mobile nodes. The values of $T_D(IM)$ and $C_P(IM)$ remain on the same order as that in Theorem 1 and 2 when ϵ is a constant.

C. Performance Analysis for Pure Unicast Traffics

Based on above discussions, we conclude that when broadcast or multicast traffics exist in the network, all of them should be utilized to generate VCC packets by performing intra-session network coding. VCC packets generated in this way can facilitate packet/service exchanges without incurring any cost. However, the situation will be complicated for scenarios with pure unicasts.

In this subsection, we assume there exist at least n unicat sessions, each of them is destined to one mobile node. We divide these unicast sessions into g groups ($1 \leq g \leq n$), each group (called *coding group*) containing n/g distinct sessions. Only the data packets destined to mobile nodes in the same group will be mixed to generate VCC packets, i.e., we perform *group-based inter-session network coding* (with *grouping parameter* g) to generate VCC packets. There may exist other unicast traffics, but they will not be involved in generating VCC packets. For pure unicast scenarios, $T_D(IM)$ is the expected delivery time for unicasting one data packet

from the infostation to a randomly selected mobile node (or in a reverse direction) under the incentive mechanism IM, and $C_P(IM)$ is the expected total number of control packets needed to support achieving $T_D(IM)$ for that data packet.

When a mobile node receives a valid VCC packet, in broadcast (or mixed traffic) scenarios this VCC packet incurs no control overhead. However, in pure unicast scenarios, this VCC packet only contains g/n useful information on average. Therefore, this VCC packet incurs $(1-g/n)$ control overhead, i.e., there is no VCC packet without incurring any cost. This is the key feature emerging in pure unicast scenarios.

Another unique feature for pure unicast scenarios is about routing. From Lemma 1 (3), we know that the network is disconnected, and because every node performs a random walk, there is no way to predict contacts. Therefore, the source or destination node cannot provide incentives to intermediate nodes except the direct next hop. As a consequence, we can only provide incentives to sustain two-hop relay [30], [31], i.e., every packet can at most take two hops from the source to the destination. In Section IV, we will remove this constraint by considering human mobility traces from the real world.

We first characterize performance of traditional approaches:

Theorem 3: [Performance of Bartering and VC for Pure Unicasts] For pure unicast traffics, (1) $T_D(B) = \Theta(n)$ and $C_P(B) = 0$; and (2) $T_D(VC) = \Theta\left(\sqrt{n}\right)$ and $C_P(VC) = \Theta\left(\sqrt{n}\right)$. (The proof is omitted here for the reason of space.)
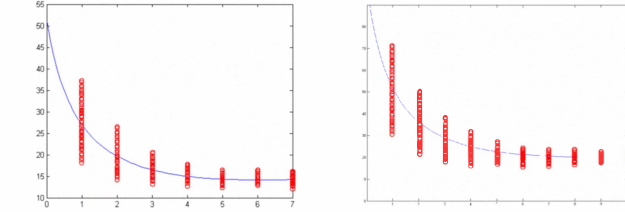
It has already been shown in [30] that when all nodes are fully cooperative, $T_D(C) = \Omega\left(\sqrt{n}\right)$ for unicasts. From Theorem 3 (2), we can see that VC based scheme can achieve the lower bound of $T_D(C)$ with large cost. This is because in order to obtain $\Theta\left(\sqrt{n}\right)$ packet delivery delay, it takes $\Theta\left(\sqrt{n}\right)$ nodes to provide packet relay services. For bartering, because there is no mutual wants between two mobile nodes at all, the only way for a mobile node to receive a packet is to contact the infostation, which causes a large delay on the order of n.

For our **C4**, by taking different values of g, i.e., the grouping parameter, **C4** in fact can provide a series of incentive mechanisms with different network performance as follows.

Theorem 4: [Performance of Our C4 for Pure Unicasts] For pure unicast traffics, (1) when $g = O\left(\sqrt{n}\right)$, $T_D(\mathbf{C4}) = \Theta\left(\sqrt{n}\right)$ and $C_P(\mathbf{C4}) = \Theta\left(\sqrt{n} - \frac{g}{\sqrt{n}}\right)$; (2) when $g = \Omega\left(\sqrt{n}\right)$, $T_D(\mathbf{C4}) = \Theta(g)$ and $C_P(\mathbf{C4}) = \Theta\left(\frac{n}{g} - 1\right)$.

Proof: (1) when $g = O\left(\sqrt{n}\right)$, from our previous work [30] we know that asking $\Theta\left(\sqrt{n}\right)$ mobile nodes to act as relay nodes is enough to achieve the lower bound of packet delivery delay on the order of $\sqrt{n}$. Asking more nodes to help only increases the cost of our **C4** and cannot further decrease the packet delivery delay. Therefore, when $g = O\left(\sqrt{n}\right)$, the infostation only randomly selects $\Theta\left(\sqrt{n}\right)$ mobile nodes from the same group to act as relay nodes. From our previous work [30], we know that in this situation $T_D(\mathbf{C4}) = \Theta\left(\sqrt{n}\right)$. Because each coded packet only contains g/n useful information for a particular mobile node, each coded packet incurs $\left(1 - \frac{g}{n}\right)$ control overhead. On average, each mobile node need exchange one coded packet with every relay node, and totally $\Theta\left(\sqrt{n}\right)$ coded packets need be exchanged for the destination node successfully decoding one data packet. Therefore, $C_P(\mathbf{C4}) = \Theta\left(\left(1 - \frac{g}{n}\right)\sqrt{n}\right)$.

(2) when $g = \Omega\left(\sqrt{n}\right)$, the infostation utilizes n/g mobile nodes (i.e., all nodes in the same group) to act as relay nodes. From the proof of Theorem 1 we know that in order to successfully decode one original data packet, on average the destination node need wait $\Theta(g)$ timeslots, and this is exactly

(a) $T_D(\mathbf{C4})$ and $C_P(\mathbf{C4})$ curve with n=50 (b) $T_D(\mathbf{C4})$ and $C_P(\mathbf{C4})$ curve with n=100

Fig. 6. **The effectiveness-cost tradeoffs of our C4 for pure unicasts.**

the $T_D(\mathbf{C4})$. Because each coded packet only contains g/n useful information for a particular mobile node, each coded packet introduces $\left(1 - \frac{g}{n}\right)$ control overhead. On average, each mobile node need exchange one coded packet with every relay node, and totally $\Theta(n/g)$ coded packets need be exchanged for the destination node successfully decoding one data packet. Therefore, $C_P(\mathbf{C4}) = \Theta\left(\left(1 - \frac{g}{n}\right)\frac{n}{g}\right)$. ∎

From Theorem 4, we directly obtain the following corollary.

Corollary 2: The effectiveness-cost tradeoff of our **C4** for pure unicasts when $C_P(\mathbf{C4}) = O(\sqrt{n})$ is given by

$$T_D(\mathbf{C4}) = \Theta\left(\frac{n}{C_P(\mathbf{C4}) + 1}\right).$$

Fig. 6 illustrate above effectiveness-cost tradeoff of our **C4** for pure unicasts. The solid lines here are theoretical results while points represent $(C_P(\mathbf{C4}), T_D(\mathbf{C4}))$ pairs obtained from simulations. For both $n = 50$ and $n = 100$, we can see that simulation results agree with our theoretical results, i.e., the effectiveness-cost tradeoff indeed exists.

From Theorem 4, Corollary 2 and simulation results, we can make the following observations about unicast traffics:

• When $C_P(\mathbf{C4}) = O(\sqrt{n})$, $T_D(\mathbf{C4})$ is a strictly decreasing function of $C_P(\mathbf{C4})$, as shown in Corollary 2 and Fig. 6. This means that there exists a fundamental tradeoff between packet delivery delay and cost. If we want to decrease packet delivery delay (i.e., increase incentive effectiveness), then the implementation cost of **C4** must be increased. By adjusting the grouping parameter, **C4** provides a systematic way to smoothly trade incentive effectiveness for implementation cost.

• In our **C4**'s tradeoff curve, when $g = n$, we obtain $C_P(\mathbf{C4}) = 0$ and $T_D(\mathbf{C4}) = \Theta(n)$. This is the same as $(C_P(B), T_D(B))$ pair (cf. Theorem 3 (1)). We note that this is not coincident because when $g = n$, no inter-session network coding is performed and no valid VCC packets generated in **C4**. So it is equivalent to bartering cases.

• In our **C4**'s tradeoff curve, when $g = 1$, we obtain $C_P(\mathbf{C4}) = \Theta(\sqrt{n})$ and $T_D(\mathbf{C4}) = \Theta(\sqrt{n})$. This is the same as $(C_P(VC), T_D(VC))$ pair (cf. Theorem 3 (2)). We note that this is also not coincident because when $g = 1$, every VCC packet only contains $1/n$ useful information, which tends to 0 when n is pretty large. Therefore, in this case, VCC packet is equivalent to e-cash packet.

IV. IMPROVING OUR C4'S PERFORMANCE WITH SOCIAL CONTACT INFORMATION

In previous section, there are two problems left for our **C4** for pure unicast scenarios, i.e., (1) two-hop relay constraints and (2) unavoidable effectiveness-cost tradeoffs. We note that these two problems are not caused by our **C4**, instead they are the consequence of oversimplified and unrealistic mobility model used for performance analysis. In this section we address these two problems by considering unique features in real-world user mobility patterns. This section also serves to validate our **C4**'s performance with human mobility traces.

TABLE IV
DATASET PROPERTIES

Dataset	Cambridge	MIT
Device	iMote	Phone
Network type	Bluetooth	Bluetooth
Duration (days)	11	246
Granularity (sec.)	600	300
# of devices	54	97
# of contacts	10,873	54,667
Data source	www.haggleproject.org	reality.media.mit.edu

In this section, we use two experimental datasets gathered by the Haggle Project (referred to as *Cambridge*) and the MIT Reality Mining Project (referred to as *MIT*). The characteristics of these datasets are summarized in Table IV. We first convert all these real life data into *social contact graphs*, with devices/mobile users as *node sets* and contacts between two nodes as *edge sets*. From these social contact graphs, we can easily find that they are *heterogeneous* both in terms of edges and community structures. Note that from the random walk mobility model, we can only obtain a homogeneous contact graph, i.e., a complete graph with equal edges. There are no link heterogeneity or community structures because every node has the same probability to contact every other node. In what follows, we show how to utilize these two kinds of heterogeneities in reality to facilitate our **C4** design. Here we *do* not mean our **C4** can work only if the information about social contact graph is available. We only want to show that **C4** can even work better if this information is available. In fact, all DTN routing schemes try to collect and utilize this information [25], and our **C4** does not require anything more.

A. Information Highway and Multi-hop Relay

The first kind of heterogeneity is called *edge heterogeneity* [32]. As shown in Fig. 7(a), edges are annotated with one or more times at which two nodes contact. We can see that not all edges in the social contact graph are with the same importance. For example, we assume that the infostation updates its status information every timeslot, and every mobile node exchanges its newest status information about the infostation with others during contacts. Then, we can observe from Fig. 7(a) that, although node G can contact the infostation, for most of timeslots, node G obtains the newest status information about the infostation from node H. So, given the existence of the information propagating highway "G−H−infostation", edge "G−infostation" can be deleted from the social contact graph. For the same reason, edge "I−infostation" and edge "G−I" can also be deleted. By continuing this deleting procedure, at last, we will obtain a tree rooted at the infostation. This tree is called the *information highway*, i.e., the structure of fast indirect paths from the infostation to all mobile nodes [32]. By assuming contacts between pairs of nodes are perfectly periodic, we can calculate the delay of the newest status updates from the infostation to every mobile node along this tree. The result is Fig. 7(b), the information highway with every node knows the latency value from the infostation to itself. The concentric circles denote ball radii increasing by 5 minutes, and the distance of each node from the common center is its latency value from the infostation.

For every node u, its latency value $d_T(u)$ is a kind of temporal distance between itself and the infostation, and therefore can be used as a routing metric. All nodes have incentives to maintain a correct $d_T(u)$ for securing its own packet delivery and the information highway can be constructed in a fully distributed way. For unicasts from mobile nodes to the infostation, multi-hop relay can be performed in the following way. When a relay node u contacts a node v, and v has a
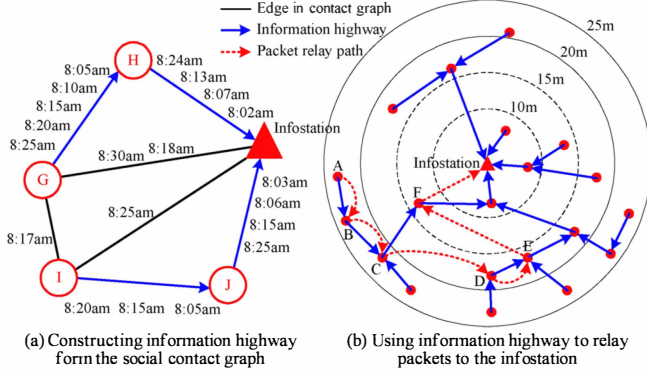
(a) Constructing information highway form the social contact graph

(b) Using information highway to relay packets to the infostation

Fig. 7. **Information highway and multi-hop relay.**

Fig. 8. **The distribution of hop counts.**

smaller $d_T(v)$, then packets destined to the infostation will be relayed to v. In Fig. 7(b), one of such relay path from node A to the infostation is illustrated by red curves. The unicasts from the infostation to the mobile nodes will be even simpler. The routing path will follow the tree of information highway, because by the definition of information highway, this is the path with the minimum packet delivery delay from the infostation to the destination.

The only problem left here is why our **C4** can sustain two-more-hop relay when the information highway emerges. For those intermediate relay nodes who do not contact with the source or the destination of packet pkt, the incentives for them to buy pkt come from the fact the the nodes which are closer to the destination will have more incentives to buy it. Therefore, they can sell it a little later and from this buy-sell procedure they can obtain more VCC packets. For example, in Fig. 7(b), on the path from node A to the infostation, node E will buy pkt from node D because he knows some nodes like E has the willingness to buy pkt, because E is closer to the infostation and has more chances to sell pkt to the infostation.

Fig. 8 shows how many hops the packets take to reach the destination. We can see that for both datasets, the hop count distributions of our **C4** are the same as that of the cases when all nodes are fully cooperative. Therefore, we can conclude that our **C4** provides adequate incentives to sustain multi-hop relay when the information highway is available.

B. Community Structure and Grouping Parameter Selection

The second kind of heterogeneity is called *community structures*. Social contact graphs typically contain parts in which the nodes are more highly connected to each other than to the rest of the graph. The set of such nodes is usually called a *community* [33]. Fig. 9(a) illustrates four communities in an exemplary social contact graph, each with a different color. We can observe that (1) a typical member in a community is linked to many other members, but not necessarily to all other members in the community, (2) different communities may overlap. In Fig. 9(a) overlapping parts are emphasized by grey color.

The use of community structures in our **C4** is straightforward. When there exists no broadcast or multicast traffics, we need to group some unicast sessions destined to different nodes to perform group based inter-session network coding to generate VCC packets. The way of grouping will significantly affect our **C4**'s performance as discussed in Section III-C. We know that if we decrease the group size (i.e., n/g) to reduce costs, the effectiveness of **C4** will also be reduced. This relationship is called effectiveness-cost tradeoffs characterized in Corollary 2. However, all these results are based on the contact graphs without community structures. For the example 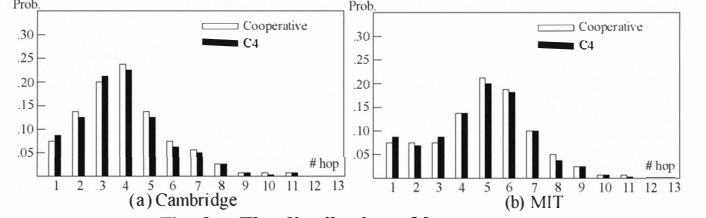given in Fig. 9(a), we can directly observe that, when we group the yellow community and blue community separately, compared to grouping yellow and blue communities together, the effectiveness will not be affected. The reason is simple: most of nodes in the yellow community in fact have no chance to contact most of nodes in the blue community. Therefore, when the VCC packets are only valid for the yellow community, there is no efficiency loss. So, every community can has its own valid VCC packets, just like every country has its own currency. Compared to a global currency, the loss of efficiency is ignorable. Therefore, when we shrink the group to the community, we can reduce overhead without losing effectiveness. Based on above discussions, we conclude that the best choice of grouping should be based on community structures in the social contact graphs.

The problem left is how to identify communities automatically. The difficulty comes from the unique features of community strictures in social graphs. Because a node in a community is not necessarily linked to all other nodes in the community, the community is not a *clique* (i.e., a complete subgraph). Different communities may overlap, and therefore traditional divisive and agglomerative methods cannot be applied [33] for this situation. Here, we use a technique called *k-clique-communities* proposed in [33]. A fully connected subgraph of k nodes is called a *k-clique*. we define a k-clique-community as the union of all k-cliques that can be reached from each other through a series of adjacent k-cliques, where two k-cliques are said to be adjacent if they share $k-1$ nodes. Fig. 9(a) illustrates overlapping 4-clique communities.

Obviously, the best choice of k here corresponds to the best choice of grouping parameter g. Here we develop a distributed scheme to find the optimal k and construct the corresponding k-clique-communities. We call the maximal complete subgraphs as *cliques*. In contrast to the k-cliques, cliques cannot be subsets of larger cliques, therefore they have to be located in a decreasing order of their size. The largest possible clique size in the graph is determined from the degree-sequence. Therefore, the infostation first finds the node with the highest degree in the social contact graph, then finds the clique containing that node. After recording this clique, the infostation deletes the node and its edges in this clique from the social contact graph. Then the infostation repeats this procedure until no nodes are left. After that, the recorded set of cliques is a set of independent k-cliques in original social contact graph. Fig. 9(b) illustrates seven independent k-cliques in Fig. 9(a), where different colors represent different k-cliques. Then the infostation chooses several largest k-cliques as *seed groups* and broadcast the node lists for each seed group. Other nodes can choose group memberships by themselves. The infostation will perform group-based inter-session network coding on packets from the same group to generate VCC packets. Since all mobile nodes have the incentives to choose the most appropriate groups for themselves, after several iterations, the best community structures will emerges by themselves and the optimal k (or g) value will be automatically determined.
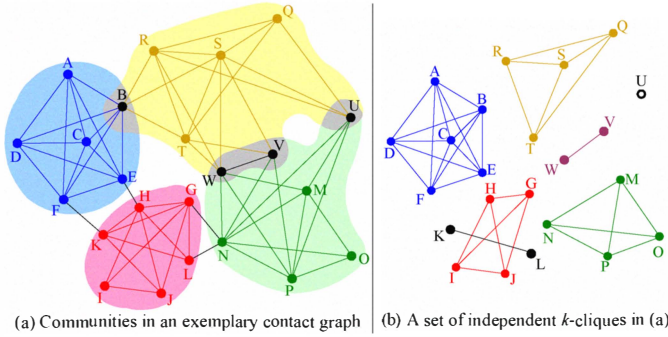
Fig. 10 shows all $(C_P(\mathbf{C4}), T_D(\mathbf{C4}))$ pairs obtained by

(a) Communities in an exemplary contact graph | (b) A set of independent k-cliques in (a)

Fig. 9. **Community structures in the social contact graph.**

Fig. 10. **The effectiveness-cost tradeoffs under different g (or k) values.**

the system itself. We can observe that there indeed exists an optimal value of g, which corresponds to the critical point at which when we further increase the cost $C_P(\mathbf{C4})$, the corresponding packet delivery delay $T_D(\mathbf{C4})$ cannot be decreased. This optimal value for Cambridge dataset is 6 while that for MIT dataset is 8. This means that, for example, the community structures of social contact graph of Cambridge dataset is most appropriately described by 6-clique-communities.

V. CONCLUDING REMARKS

The benefits of utilizing network coding in wireless networks, such as improving throughput, reducing energy consumption, simplifying transmission scheduling, etc., have been extensively studied in the literature (cf. [26], [27], [30] and references therein). However, as far as we know, our **C4** is the first to demonstrate a new potential benefit brought by network coding for wireless networks, i.e., providing incentives for cooperation. Incentive problems in wireless networks with network coding have also been studied in [23], [24]. Both of them apply game theory to study coding strategies adopted by individual nodes and still use virtual currency (credit) to create incentives. In our **C4**, we assume that the main application of MWNs is to provide Internet access services, and most traffic in a MWN will go through the infostation, which enforces random linear network coding (cf. Section II-A) to improve the whole system's efficiency. Therefore, no room is left for self-interested mobile nodes to change coding strategies, and coded packets can be safely utilized to create incentives. By treating network coding as a tool to produce virtual commodity currency, our **C4** provides an effective and lightweight solution to induce cooperation, which is impossible for [23], [24] based on traditional virtual currency. How to extend our **C4** to pure wireless ad hoc networks will be an interesting and important topic and we will study it in our future work.

REFERENCES

[1] L. Buttyan and J.-P. Hubaux, *Security and Cooperation in Wireless Networks*. Cambridge, UK: Cambridge University Press, 2007.

[2] F. S. Mishkin, *The Economics of Money, Banking and Financial Markets (7th edn.)*. Reading, MA: Addison-Wesley Pub. Co., 2003.

[3] R. Davies, *A History of Money from Ancient Times to the Present Day (3rd edn.)*. Wales, UK: University of Wales Press, 2002.

[4] W. H. Yuen, R. D. Yates, and S.-C. Mau, "Exploiting data diversity and multiuser diversity in noncooperative mobile infostation networks," in *Proc. of InfoCom 2003*, San Francisco, CA, March 2003.

[5] P. Ganesan and M. Seshadri, "On cooperative content distribution and the price of barter," in *Proc. of ICDCS 2005*, Columbus, OH, June 2005.

[6] L. Buttyan, L. Dora, M. Felegyhazi, and I. Vajda, "Barter trade improves message delivery in opportunistic networks," *Ad Hoc Networks*, vol. 8, no. 1, pp. 1–14, January 2010.

[7] F. Milan, J. J. Jaramillo, and R. Srikant, "Achieving cooperation in multihop wireless networks of selfish nodes," in *Proc. of 2006 workshop on Game theory for communications and networks*, Pisa, Italy, Oct. 2006.

[8] U. Shevade, H. H. Song, L. Qiu, and Y. Zhang, "Incentive-aware routing in DTNs," in *Proc. of IEEE ICNP 2008*, Orlando, FL, October 2008.

[9] Z. Liu, H. Hu, Y. Liu, K. Ross, Y. Wang, and M. Mobius, "P2P trading in social networks: The value of staying connected," in *Proc. of InfoCom 2010*, San Diego, CA, March 2010.
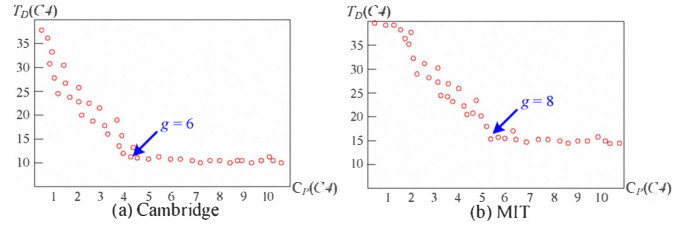
[10] S. Zhong, J. Chen, and Y. R. Yang, "Sprite: A simple, cheat-proof, credit-based system for mobile ad-hoc networks," in *Proc. of InfoCom 2003*, San Francisco, CA, March 2003.

[11] J. Crowcroft, R. Gibbens, F. Kelly, and S. Ostring, "Modelling incentives for collaboration in mobile ad hoc networks," *Performance Evaluation*, vol. 57, no. 4, pp. 427–439, August 2004.

[12] N. B. Salem, L. Buttyan, J. Hubaux, and M. Jakobsson, "Node cooperation in hybrid ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 4, pp. 1–12, April 2006.

[13] W. Wang, S. Eidenbenz, Y. Wang, and X.-Y. Li, "OURS: optimal unicast routing systems in non-cooperative wireless networks," in *Proc. of MobiCom 2006*, Los Angeles, CA, Sept. 2006.

[14] S. Zhong and F. Wu, "On designing collusion-resistant routing schemes for non-cooperative wireless ad hoc networks," in *Proc. of MobiCom 2007*, Montreal, Quebec, Canada, Sept. 2007.

[15] V. Srinivasan, P. Nuggehalli, C.-F. Chiasserini, and R. R. Rao, "Cooperation in wireless ad hoc networks," in *Proc. of InfoCom 2003*, San Francisco, CA, March 2003.

[16] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," in *Proc. of MobiCom 2000*, Boston, MA, August 2000.

[17] S. Buchegger and J.-Y. L. Boudec, "Performance analysis of the CONFIDANT protocol," in *Proc. of MobiHoc 2002*, Lausanne, Switzerland, June 2002.

[18] J. Jaramillo and R. Srikant, "DARWIN: Distributed and adaptive reputation mechanism for wireless networks," in *Proc. of MobiCom 2007*, Montreal, Quebec, Canada, Sept. 2007.

[19] Y.-D. Lin and Y.-C. Hsu, "Multihop cellular: a new architecture for wireless communications," in *Proc. of InfoCom 2000*, Tel Aviv, Israel, March 2000.

[20] I. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: a survey," *Computer Networks*, vol. 47, no. 4, pp. 445–487, March 2005.

[21] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc. of InfoCom 2009*, Rio de Janeiro, Brazil, April 2009.

[22] H. Wu, R. Fujimoto, R. Guensler, and M. Hunter, "MDDV: a mobility-centric data dissemination algorithm for vehicular networks," in *Proc. of ACM VANET 2004*, Philadelphia, PA, Oct. 2004.

[23] X. Zhang and B. Li, "DICE: a game theoretic framework for wireless multipath network coding," in *Proc. of MobiHoc 2008*, Hong Kong, China, May 2008.

[24] T. Chen and S. Zhong, "INPAC: An enforceable incentive scheme for wireless networks using network coding," in *Proc. of InfoCom 2010*, San Diego, CA, March 2010.

[25] Z. Zhang, "Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: Overview and challenges," *IEEE Communications Surveys and Tutorials*, vol. 8, no. 1, pp. 24–37, March 2006.

[26] T. Ho, R. Koetter, M. Medard, D. R. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," in *Proc. of ISIT 2003*, Yokohama, Japan, June-July 2003.

[27] S. Deb, M. Medard, and C. Choute, "Algebraic gossip: A network coding approach to optimal multiple rumor mongering," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2486–2507, June 2006.

[28] C. Ortolf, C. Schindelhauer, and A. Vater, "Paircoding: Improving file sharing using sparse network codes," in *Proc. of ICIW 2009*, Venice, Italy, May 2009.

[29] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[30] C. Zhang, X. Zhu, and Y. Fang, "Throughput-delay tradeoffs in large-scale manets with network coding," in *Proc. of InfoCom 2009*, Rio de Janeiro, Brazil, April 2009.

[31] M. Grossglauser and M. D. Tse, "Mobility increases the capacity of ad hoc wireless networks," in *Proc. of InfoCom 2001*, Anchorage, Alaska, April 2001.

[32] G. Kossinets, J. Kleinberg, and D. Watts, "The structure of information pathways in a social communication network," in *Proc. of KDD 2008*, Las Vegas, NE, August 2008.

[33] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, pp. 814–818, June 2005.