ARTICLE IN PRESS

Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Immunization-based redundancy elimination in Mobile Opportunistic Networks-Generated big data

Junbao Zhang [a], Haojun Huang [b,*], Yan Luo [c], Yinting Fan [a], Guan Yang [a]

[a] School of Computer Science, Zhongyuan University of Technology, Zhengzhou, 450007, China
[b] Electronic Information School, Wuhan University, Wuhan, 430072, China
[c] Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, USA

ARTICLE INFO

Article history:
Received 28 May 2017
Received in revised form 15 August 2017
Accepted 24 August 2017
Available online xxxx

Keywords:
Big data
Redundancy elimination
Immunization
Multi-copy

ABSTRACT

Diverse sensors and smart devices are promising in facilitating to perform specific tasks which generate massive data whilst such data transmission is challenged in the big data era. Under some circumstances, these devices may form Mobile Opportunistic Networks (MONs) which are characterized by intermittent connectivity. In such scenarios, there is a critical issue that nodes with the already delivered message copies may continue carrying and transmitting the copies if they are not informed that the message has been delivered. This may result in redundant data, and thus large consumption of network resources and network performance degradation. To avoid generating, transmitting and storing unwanted data due to redundant message copies, we propose an Immunization-Based Redundancy Elimination scheme (IBRE) in MONs to stop useless data transmission and flush redundancy. In IBRE, each destination independently selects the right number of ACKs distributed to respond to the variation of the amount of redundant data in a dynamic fashion. Simulation results demonstrate that IBRE suppresses redundant data transmission and eliminates useless data generated from redundant message copies in cost-efficient manner.

1. Introduction

Over the past few years, big data has attracted much attention in both academia and industry. With the proliferation of sensors and smart devices, massive data has been collected to perform specific tasks through various networks, such as Internet of Things (IoT) [1,2], Software-Defined Networking (SDN) [3] and Mobile Opportunistic Networks (MONs). Among of them, MONs, as a promising network paradigm, experience intermittent connectivity [4] owing to node mobility, energy conservation, scheduling, and so on, but can provide ubiquitous communication for end-users. Due to these inherent characteristics of MONs, various MON applications, like environment monitoring, wildlife tracking, disaster relief, and tactical communication [5,6], have been developed by exploiting multi-copy routing [4]. Taking into account the fact that the multi-copy routing protocols cause unwanted redundant data for MONs, in this paper, we focus on the immunization-based redundancy elimination in MONs-Generated big data.

Prior studies have proposed numerous schemes to transmit data in MON scenarios from different perspectives, like flooding, one-hop, direct delivery, etc. Amongst these schemes, multi-copy routing is a promising solution [7–9] which distributes duplicate copies per message to all nodes or a subset of them in the network. After a message was generated, some nodes may carry copies of the message over time. Apparently this can improve the chance of encountering the destination of the message. Therefore high probability of delivery and small delays may be achieved. However, multiple message copies may also lead to resource contention. Although multi-copy routing protocols can improve delivery ratio and reduce delivery delay, the performance of them may be effected by the congestion.

Moreover, after the destination receives a message, copies of the message carried by the source node and relay nodes are useless, i.e. redundant. If the source and relays are not informed that the message has been delivered, they may still keep the replicas of the message. Namely, they would continue carrying and forwarding copies of the already delivered message. Such redundant data may occupy the buffers, and be transmitted to other nodes sequentially. Thereby, redundant data is extremely wasteful of resources, like buffer space, wireless bandwidth and energy. Furthermore, useful data cannot be efficiently transmitted since the performance of network degrades significantly due to the unwanted data.

As a result of the intermittent connectivity of MONs, it is hard to suppress unwanted data transmission and eliminate redundant data. **Redundancy Elimination** is a critical issue in MONs [10–12] and deserves close attention. The main concern of it is how

* Corresponding author.
E-mail addresses: junbao.zhang@zut.edu.cn (J. Zhang), hjhuang@whu.eud.cn (H. Huang), YanLuo@uml.edu (Y. Luo), fytzcx@aliyun.com (Y. Fan), guanyang@zut.edu.cn (G. Yang).

http://dx.doi.org/10.1016/j.future.2017.08.059
0167-739X/

to respond to the variation of the amount of redundant data. To address this issue, we propose a novel redundancy elimination scheme in which each destination independently selects the right number of ACK responding to the number of infected nodes in a dynamic way. We also conduct extensive simulations to evaluate the performance of IBRE regarding to suppressing redundant data transmission and eliminating useless data.

The main contributions of our work are summarized as follows:

(1) We present and discuss issues of redundancy elimination including motivation and classification of existing schemes.
(2) We formally analyze the expected number of infected nodes that an immune node may encounter.
(3) We propose a novel redundancy elimination scheme IBRE to stop redundant data transmission and eliminate such redundant data. In IBRE, each destination node independently selects the right number of ACKs disseminated in a dynamic manner to adapt to the variation of the amount of redundant data.
(4) We conducted extensive simulations to evaluate the proposed scheme with real traces. Simulation results show the benefit of redundancy elimination for MONs.

The remainder of this paper is organized as follows. In the next section, we motivate and formulate the issues of redundancy elimination. Section 3 describes our proposed scheme, IBRE, in detail. Simulation results are presented in Section 4. Finally, we summarize our work in Section 5.

2. Redundancy elimination

In this section, we explore the issues of redundancy elimination in MONs in which some redundant data may be generated from useless message copies, transmitted to other uninfected nodes and stored in the buffer of infected nodes. After that, we also review some existing redundancy elimination schemes.

Assume that each message has a field of the replication value and a multi-copy routing protocol is utilized in a MON. Given a message m, let us call a node **"infected"** when it has one or more copies[1] of m, and **"uninfected"** otherwise. And we call a node **"immune"** when it has the corresponding ACK for m.

2.1. Motivation

Multi-copy routing protocols distribute multiple copies of a message in the network. Whenever a source node creates a message m, it keeps the message until a contact opportunity arises. If the source encounters another node, whether it disseminates the message copy or not depends on the forwarding strategy. Some copies of message m are spread over time, namely some nodes including the source and relays are infected. The number of message copies (equivalently the number of infected nodes) hinges on the routing protocol used and mobility pattern of nodes.

Maybe at a certain instant, the destination node d receives a copy of message m. Since then replicas of message m carried by infected nodes become redundant. Nevertheless infected nodes including the source and relays may still keep the redundant copies if they are not informed that the message has been delivered. In other words, without a redundancy elimination scheme, the infected nodes would continue carrying and forwarding the redundant message copies to uninfected nodes [13]. Redundancy may sequentially occupy limited buffers and waste resources, like energy, wireless bandwidth and memory space. This may potentially result in extremely resource waste. Furthermore, redundant message copies may degrade the performance of multi-copy routing protocols. That is why the redundant message replicas must be eliminated from the network.

Redundancy elimination is a critical concern for multi-copy routing protocols in MONs. The natural objective of the redundancy elimination scheme is to stop useless data transmission and eliminate replicas of already delivered messages from the network, and impose no or little overhead [10]. To achieve this, some redundancy elimination scheme have been proposed, also called recovery schemes [14], such as VACCINE [10], Intermediate Immunity [15], threshold-based [16], etc.

2.2. Classification

According to the characteristic relevant to whether the scheme uses ACK (acknowledgment message, or anti-packet is also used [10]) or not, we categorize existing redundancy elimination schemes into two types [17]: ACK-based (i.e. explicit notification) and non-ACK (i.e. implicit notification).

Non-ACK schemes are based on a threshold without an ACK, like JUST_TTL [10], Threshold-based [16]. JUST_TTL is the basic form of Non-ACK scheme, in which if it is expired, namely the value of Time To Live (TTL) field is equal to 0, the message will be dropped. Threshold-based [16] sets a threshold for each message. If the threshold is reached, the infected node will clean up the message copy from its buffer. This type of schemes is simple, and can reduce resource waste. Yet it is difficult to set an appropriate threshold in MON scenarios. Owing to the frequently changing of network topology in MONs, non-ACK schemes often suffer from the threshold selection. Furthermore, our extensive simulation results also show that the performance of a routing protocol is often beset by the threshold-based redundant elimination scheme utilized.

ACK-based schemes utilize ACKs to notify infected nodes of the delivery [10]. ACK is used as a special type of control message to acknowledge infected nodes that the corresponding message has been delivered. Each ACK has a field associated with a specific message identifier. Different from data message, the size of ACK is fixed and small. That would not violate the design goal of redundancy elimination. ACK-based schemes usually consist of two parts: (i) notification: as the destination first receives a message, it generates a corresponding ACK and distributes the ACK to the network; (ii) elimination: nodes receiving an ACK will not receive the corresponding message again, no matter whether they are infected or not. If they are infected, they also drop the message copies from their buffer.

VACCINE [10] and Broadcast-based [13] are two typical ACK-based schemes. After a message achieves its destination, the former floods ACKs in the network, the latter broadcasts ACKs over the more powerful radio which is unrealistic in some scenarios like a network composed of nodes equipped with hand-hold devices. Although ACK-based schemes can effectively eliminate some redundant message copies, it is a tradeoff between performance and the number of ACKs distributed in the network. The main concern of ACK-based schemes is how to choose the right number of ACKs. In addition, Gao et al. [18] investigated characteristics of forwarding redundancy from both theory and experiment respectively. Based on that, they proposed adaptive schemes to control and eliminate the forwarding redundancy in distributed manner.

In this paper, we prefer the ACK-based redundant elimination scheme. In light of dynamic nature of MONs, flooding or limiting a fixed quota of ACKs may not be appropriate to respond to the current conditions of the network. To this end, we propose a novel redundant elimination scheme which will be presented in the next section.

[1] Like spray and wait [7], if the replication value of a message is greater than 1, we say the node carrying the message has "more copies".