Ad Hoc Networks 40 (2016) 26–36

Contents lists available at ScienceDirect

Ad Hoc Networks

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

Using data mules for sensor network data recovery

Jon Crowcroft[a], Liron Levin[b], Michael Segal[b,*]

[a] Computer Laboratory, Cambridge University, Cambridge, United Kingdom
[b] Communication Systems Engineering Department, Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Article history:
Received 20 March 2015
Revised 21 November 2015
Accepted 27 December 2015
Available online 4 January 2016

Keywords:
Data mule
Data recovery
Sensor networks

ABSTRACT

In this paper, we study the problem of efficient data recovery using the data mules approach, where a set of mobile sensors with advanced mobility capabilities re-acquire lost data by visiting the neighbors of failed sensors, thereby avoiding permanent data loss in the network. Our approach involves defining the optimal communication graph and mules' placements such that the overall traveling time and distance is minimized regardless to which sensors crashed. We explore this problem under different practical network topologies such as arbitrary graphs, grids and random linear networks and provide approximation algorithms based on multiple combinatorial techniques. Simulation experiments demonstrate that our algorithms outperform various competitive solutions for different network models, and that they are applicable for practical scenarios.

1. Problem formulation

A data mule is a vehicle that physically carries a computer with storage between remote locations to effectively create a data communication link [21]. In ad-hoc networks, data mules are usually used for data collection [5] or monitoring purposes [11] when the network topology is sparse or when communication ability is limited. In this paper, we propose to extend the usage of data mules to the critical task of network reliability. That is, using the advantages of mobility capabilities to prevent losing crucial information while taking into consideration the additional operational costs. We propose to model the penalty of a sensor crash as the cost of restoring its information loss, and present several algorithms that minimize the total cost given any combination of failures. We use concepts from graph theory to model the deployment of the ad-hoc network and give special attention to linear and grid graph models, whose unique network characteristics makes them well suited for many sensor applications such as monitoring of international borders, roads, rivers, as well as oil, gas, and water pipeline infrastructures [11,13].

Let T be a data gathering tree rooted at root r spanning n wireless sensors positioned in the Euclidean plane, where data propagates from leaf nodes to r. We model the environment as a complete directed graph $G=(V,E)$, where the node set represents the wireless sensors and the edge represents distance or time to travel between that sensors. We assume the sensors are deployed in rough geographic terrain with severe climatic conditions, which may cause sporadic failures of sensors. Clearly, if a sensor v fails, it is undesirable to lose the data it collected from its children in T, $\delta(v,T)$. Thus, a group of data gathering mules must travel through $\delta(v,T)$ and restore the lost information. We define this problem as (α, β)-Mule problem, where α is the number of simultaneous node failures and β is the number of traveling mules.

For $\alpha=1, \beta=1$, the mule visits the children of v over the shortest tour, $t(m, \delta(v,T))$, starting and ending at node $m \in V$, where the length of the tour is equal to the Euclidean length of distances; the goal is to find a data gathering tree T, the placement of the mule m, and the shortest tours, $t(m, \delta(v,T))$ for all $v \in V$, which minimize the total

* Corresponding author. Tel.: +972 8 6477234; fax: +972 8 6472883.
E-mail addresses: Jon.Crowcroft@cl.cam.ac.uk (J. Crowcroft), levinlir@bgu.ac.il (L. Levin), segal@bgu.ac.il, segal@cse.bgu.ac.il, michael99segal@gmail.com (M. Segal).

http://dx.doi.org/10.1016/j.adhoc.2015.12.009
1570-8705/

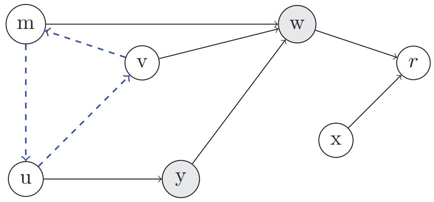

Fig. 1. Example for the mule tour when 2 nodes fail. The grey nodes represent sensors that experienced failure and the blue dashed lines represent the mule tour; the tour starts and ends at node m. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

traveling distance given **any** sensor can fail. Formally, find T and m such that $\sum_{v\in V} |t(m, \delta(v, T))|$ is minimized. In a similar way, we can define the problem for $\alpha > 1, \beta = 1$ (see example for $\alpha = 2$ in Fig. 1, where the edges are directed towards the root). Formally, find T and m such that $\sum_{\{F\subset V:|F|=\alpha\}} |t(m, \bigcup_{v\in F} \delta(v, T))|$ is minimized. We can extend this scenario to the case where instead of a single mule, we have β mules $\bar{m} = \{m_1, m_2, \dots, m_\beta\}$ deployed at different coordinates of the graph. When a node fails, its children must be visited by **one** of the mules to restore the lost data, which can be viewed as a mule assignment per node for the single node failure, or per unique node failure combination for the multi-failures case. In addition to T, we must find the location of all mules $\bar{m}$, and an assignment of each node $v \in V$ to a mule $m_i \in \bar{m}$ that minimizes the total travel cost of all mules. Formally, for $\beta > 1$, let $t(m_i, \delta(v, T))$ be the shortest path tour that includes mule m_i and the children of node v that mule m_i should visit. For $\alpha = 1$, the optimization problem is to find T and m such that $\sum_{v\in V} \sum_{m_i\in\bar{m}} |t(m_i, \delta(v, T))|$ is minimized.

We consider two network models, *complete* graphs and *unit disc* graphs. In the complete graph model, there is a directed edge between any pair of nodes in the graphs while in the unit disc graph model, there is an edge if and only if $d(u, v) \le 1$, where $d(u, v)$ is the Euclidean distance between nodes u and v.

A summary of symbols used throughout this papers are depicted in Table 1.

1.1. Our contribution

To the best of our knowledge, this is the first work exploring the mule approach for avoiding data loss due to communication failures. Our results are summarized in the following (Table 2):

Table 2
Summary of results.

Underlying graph	Problem	Topology	Approximation ratio
Complete	(1, 1)-Mule	Arbitrary	$1 + 1/c, c > 1$
	(α, 1)-Mule		$\min(3, 1 + s^*)$, $s^* = \min_{v\in V} \frac{\max d(v,u)}{\min d(v,w)}$
	(1, β)-Mule		2
UDG	(1, 1)-Mule	Line	OPT
	(α, 1)-Mule	Line	OPT
	(1, 1)-Mule	Random Line	4
	(1, 1)-Mule	Grid	$1 + (2 + \sqrt{2})/\sqrt{n}$

1.2. Paper outline

The paper is organized as follows. In the next section we discuss the previous related work to our problem. We analyze different variations of the mule problem under the complete graph model and the unit disc graph model in Sections 3 and 4, respectively. Section 5 outlines a possible distributed implementation of our algorithms. In Section 6 we present simulations of our algorithms under different network settings and conclude in Section 7.

2. Related work

Exploiting mobile data carriers (mules) in ad-hoc networks has received significant attention recently. The subject of major interest in most works is using the mules to relay and collect messages in sparse network settings, where adjacent sensors are far from each other, in order to substantially reduce the cost of indeterminate sensors communication and data aggregation. For example, Wu et al. [22], investigate how to use the mule architecture to minimize data collection latency in wireless sensor networks. They reduce this problem to the well-known k-traveling salesperson with neighborhood and provide a constant approximation algorithm and two heuristic for it. In a related paper by Ciullo et al. [8], the collector is responsible for gathering data messages by choosing the optimal path that minimizes the total transmitted energy of all sensors subject to a maximum travel delay constraint. In their model, each sensor sends different amount of data. The authors also use the k-traveling salesperson with neighborhood problem in their solution technique and prove both analytically and through simulation that letting the mobile collector come closer to sensors with more data to transmit leads to significant reduction in energy consumption. Cheong et al. [6] investigate how to find a data collection path for a mobile base

Table 1
Symbol table.

m	The mule placement in T
$\delta(v, T)$	The children of node v in tree T.
$\|t(m, \delta(v, T))\|$	The cost of the shortest tour visiting the children of node v in tree T starting from node m.
$c(m, r)$	Total cost of the data gathering tree when mule is placed at node m and root is placed at node r. The notation is used for topologies for which the cost of the solution solely depends on m and r.
$\pi(i, m, r)$	Number of times node i is visited by the mule for a given m and r.
$c(T)$	The cost of a tree solution T when the placement of m and r is given in advance.