Optimizing Information Credibility in Social Swarming Applications

Bin Liu*, Peter Terlecky‡, Amotz Bar-Noy‡, Ramesh Govindan*, Michael J. Neely*
*University of Southern California, Los Angeles ‡City University of New York

***Abstract*—With the advent of smartphone technology, it has become possible to conceive of entirely new classes of applications. *Social swarming*, in which users armed with smartphones are directed by a central *director* to report on events in the physical world, has several real-world applications. In this paper, we focus on the following problem: how does the director optimize the selection of reporters to deliver credible *corroborating* information about an event? We first propose a model, based on common intuitions of believability, about the credibility of information. We then cast the problem as a discrete optimization problem, and introduce optimal centralized solutions and an approximate solution amenable to decentralized implementation whose performance is about 20% off on average from the optimal while being 3 orders of magnitude more computationally efficient. More interesting, a time-averaged version of the problem is amenable to a novel stochastic utility optimization formulation, and can be solved optimally, while in some cases yielding decentralized solutions.**

I. Introduction

With the advent of smartphone technology, it has become possible to conceive of entirely new classes of applications. Recent research has considered personal reflection [1], social sensing [2], lifestyle and activity detection [3], and advanced speech and image processing applications [4]. These applications are enabled by the programmability of smartphones, their considerable computing power, and the presence of a variety of sensors on-board.

In this paper, we consider a complementary class of potential applications, enabled by the same capabilities, that we call *social swarming*. In this paper, we consider a constrained form of a social swarming application in which N participants, whom we call *reporters*, collaboratively engage in a well-defined task. Each reporter is equipped with a smartphone and directly reports to a swarm director using the 3G/EDGE network. A reporter may either be a human being or a sensor (static, such as a fixed camera, or mobile, as a robot). A *director* (either a human being, or analytic software) assimilates these reports, and may perform some actions based on the content of these combined reports.

Each reporter reports on an *event*. The nature of the event depends upon the social swarming application: for example, in a search and rescue operation, an event corresponds to the sighting of an individual who needs to be rescued; in the balloon hunt, an event is the sighting of a balloon[1]. Events occur at a particular *location*, and multiple events may occur concurrently either at the same location or at different locations.

Reporters can transmit reports of an event using one of several formats: as a video clip, an audio clip, or a text message describing what the report sees. Each report is a form of *evidence* for the existence of the event. In general, we assume that each reporter is capable of generating R different report formats, denoted by f_j, for $1 \leq j \leq R$. However, different formats have different costs to the network: for example, video or audio consumes significantly higher transmission resources than text. We denote by e_j the cost of a report f_j: for ease of exposition, we assume that reports are a fixed size so that all reports of a certain format have the same cost (our results can be easily generalized to the case where report costs are proportional to their length). Finally, reporters can be mobile, but we assume that the director is aware of the location of each reporter.

Now, suppose that the director in a swarming application has heard, through out-of-band channels or from a single reporter, of the existence of an event E at location L. To verify this report, the director would like to request *corroborating* reports from other reporters in the vicinity of L. Which reporters should she get corroborating reports from? What formats should those reporters use?

To understand this, recall that the goal of corroboration is to increase the director's belief in the occurrence of the event. This depends upon the credibility of the report, which we model using two common intuitions about credibility. The first intuition is based on the maxim "seeing is believing": a video report is more credible than a text report. We extend this maxim in our model to incorporate other formats, like audio: audio is generally less credible than video (because, while it gives some context about an event, video contains more context), but more credible than text (for a similar reason).

Our second intuition is based on the often heard statement "I'll believe someone who was there", suggesting that proximity of the reporter to an event increases the credibility of the report. More precisely, a report A generated by a reporter at distance d_a from an event has a higher credibility than a report B generated by a reporter at a distance d_b, if $d_a < d_b$.

More formally, let S_i be the position of reporter i, L be the position of event E and $c_{i,j}(S_i, L)$ be the credibility of the

This research was sponsored by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

[1] http://www.crn.com/networking/222000334

TABLE I
NOTATION

N	the total number of available reporters
$c_{i,j}$	credibility of reporter i using format j (see (1))
R	the total number of report formats
e_j	the cost when using report format f_j
C	the target credibility in MinCost
A	the dynamic programming process of MinCost
B	the cost budget in MaxCred
D	the dynamic programming process of MaxCred

report generated by reporter i when report format f_j is used. Then, a $c_{i,j}(S_i, L)$ defined as:

$$c_{i,j}(S_i, L) = \begin{cases} \gamma_j / d(S_i, L)^{\delta_j}, & \text{if } h_0 < d(S_i, L) \\ \gamma_j / h_0^{\delta_j}, & \text{if } d(S_i, L) \le h_0 \end{cases} \tag{1}$$

with $1 \le j \le R$, $\gamma_1 \le \gamma_2 \le \cdots \le \gamma_R$, and $\delta_1 > \delta_2 > \cdots > \delta_R$, incorporates the intuitions described above. Here, $d(.)$ is the Euclidean distance between points, h_0 is a certain minimum distance to avoid division by zero as well as to bound the maximum credibility to a certain level, γ_j is a constant of proportionality implying the maximum achievable credibility of report format f_j, and the credibility decays according to a power-law with exponent δ_j when format f_j is used.

Although we have assigned objective quantitative values to credibility, belief or disbelief is often qualitative and subjective. Thus, we expect swarm directors to operate in one of two modes: *a*) ask the network to deliver corroborating reports whose total credibility is above a certain threshold, while minimizing cost, or *b*) obtain as much corroborating information that they can get from the network for a given cost. We study these two formulations, respectively called MinCost and MaxCred. Our formulations use an *additive corroboration function* which defines total credibility as the sum of individual credibilities. Moreover, the intuition for a particular threshold value C can be explained as follows. Suppose a director would be subjectively satisfied with 3 corroborating video clips from someone within 10m of an event. One could translate this subjective specification into a threshold value by simply taking the sum of the credibilities of 3 video reports from a distance of 10m.

In the next two sections, we formally define MinCost and MaxCred, and then consider two problem variants: a *one-shot* problem which seeks to optimize reporting for individual events, and a *renewals* problem which optimizes reporting over a sequence of event arrivals. The introduction of a novel problem setting and an exploration of the one-shot and the renewals problem are the main contributions of the paper.

II. The One-Shot Problem

In this section, we formally state the MinCost and MaxCred formulations for the additive corroboration function and develop optimal solutions for them, and then explore an approximation algorithm that leverages the structure of the credibility function for efficiency. Our exposition follows the notation developed in the previous section, and summarized in Table I.

A. Formulation and Complexity of MinCost and MaxCred:

1) Problem Formulations: In the previous section, we informally defined the MinCost problem to be: what is the minimum cost that guarantees total credibility $C > 0$? MinCost can be stated formally as an optimization problem:

$$\begin{aligned} \text{Minimize :} \quad & \textstyle\sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,j} e_j \\ \text{Subject to:} \quad & \textstyle\sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,j} c_{i,j} \ge C \\ & x_{i,j} \in \{0,1\}, \forall i \in \{1, ..., N\}, \forall j \in \{1, ..., R\} \\ & \textstyle\sum_{j=1}^{R} x_{i,j} \le 1, \forall i \in \{1, ..., N\} \end{aligned} \tag{2}$$

where $x_{i,j}$ is a binary variable that is 1 if reporter i uses format f_j, and 0 otherwise.

Analogously, we can formulate MaxCred (the maximum credibility that can be achieved for a cost budget of $B > 0$) as the following optimization problem:

$$\begin{aligned} \text{Maximize :} \quad & \textstyle\sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,j} c_{i,j} \\ \text{Subject to:} \quad & \textstyle\sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,j} e_j \le B \\ & x_{i,j} \in \{0,1\}, \forall i \in \{1, ..., N\}, \forall j \in \{1, ..., R\} \\ & \textstyle\sum_{j=1}^{R} x_{i,j} \le 1, \forall i \in \{1, ..., N\} \end{aligned} \tag{3}$$

2) On the Complexity of MinCost *and* MaxCred*:* If, in the above formulation, the cost e_j is also dependent on the identity of the reporter (and therefore denoted by $e_{i,j}$), the MaxCred problem generalizes to the Multiple-Choice Knapsack Problem (MCKP, [7]). Moreover, the special case of one format (and $e_{i,j} = e_i$) is the well-known Knapsack problem (KP) which is NP-hard. However, when the cost is dependent only on the format (i.e., $e_{i,j} = e_j$), we can state the following theorem, whose proof (omitted for brevity) uses a reduction from the original Knapsack problem.

Theorem 2.1: MinCost and MaxCred are NP-Hard.

B. Optimal Solutions

Despite Theorem 2.1, it is instructive to consider optimal solutions for the two problems for two reasons. First, for many social swarming problem instances, the problem sizes may be small enough that optimal solutions might apply. Second, optimal solutions can be used to calibrate an approximation algorithm that we discuss later. In this section, we discuss the *dynamic programming* based optimal solutions.

Since there exist optimal, weakly-polynomial algorithms for MCKP, it is natural that similar algorithms exist for MinCost and MaxCred. We describe these algorithms for completeness, since we use them in a later evaluation.

For MinCost (2), we can write $y_{i,j} = 1 - x_{i,j}$, where $y_{i,j} \in \{0, 1\}$, and then we have:

$$\begin{aligned} & N \textstyle\sum_{j=1}^{R} e_j - \text{Maximize: } \sum_{i=1}^{N} \sum_{j=1}^{R} y_{i,j} e_j \\ \text{Subject to:} \quad & \textstyle\sum_{i=1}^{N} \sum_{j=1}^{R} y_{i,j} c_{i,j} \le \sum_{i=1}^{N} \sum_{j=1}^{R} c_{i,j} - C = W \\ & y_{i,j} \in \{0,1\}, \forall i \in \{1, \ldots, N\}, \forall j \in \{1, \ldots, R\} \\ & \textstyle\sum_{j=1}^{R} y_{i,j} \ge R - 1, \forall i \in \{1, \ldots, N\} \end{aligned} \tag{4}$$

where the minimization problem (2) has been transformed into a maximization problem, and the notation in (4) emphasizes that the first term in the total cost $N \sum_{j=1}^{R} e_j$ does not depend on the $y_{i,j}$ variables to be optimized. For a given event, the sum of the $c_{i,j}$ values is a constant, and so W is also a constant.

This optimization problem can be solved by a dynamic programming approach if we assume all $c_{i,j}$s are truncated to a certain decimal precision, so that $c_{i,j} \in \{0, \zeta, 2\zeta, \ldots\}$

where ζ is a discretization unit. Then for any binary $y_{i,j}$ values that meet the constraints of the above problem, the sum $\sum_{i=1}^{N}\sum_{j=1}^{R} y_{i,j}c_{i,j}$ takes values in a set $\mathcal{W}\triangleq\{0,\zeta,2\zeta,\ldots,W\}$. Note that the cardinality $|\mathcal{W}|$ depends on N, R, the $c_{i,j}$ values, and the discretization unit ζ. Now define $A(l,s)$ as the sub-problem of selecting reporters in the set $\{1,\ldots,l\}$ subject to a constraint s. Assuming $A(l,s)$ values are known for a particular l, we recursively compute $A(l+1,s)$ for all $s\in\mathcal{W}$ by:

$$A(l+1,s) = \max[\phi^{(0)}(l,s),\phi^{(1)}(l,s),\ldots,\phi^{(R)}(l,s)] \qquad (5)$$

where $\phi^{(k)}(l,s)$ is defined for $k\in\{0,1,\ldots,R\}$:

$$\phi^{(k)}(l,s)\triangleq A(l,s-\textstyle\sum_{j=1,j\neq k}^{R} c_{l,j}) + \sum_{j=1,j\neq k}^{R} e_j$$

This can be understood as follows: The value $\phi^{(k)}(l,s)$ is the cost associated with reporter $l+1$ using option $k\in\{0,1,\ldots,R\}$ and then allocating reporters $\{1,\ldots,l\}$ according to the optimal solution $A(l,s-\sum_{j=1,j\neq k}^{R} c_{l,j})$ that corresponds to a smaller budget. Note that option $k\in\{1,\ldots,R\}$ corresponds to reporter $l+1$ using a particular format (so that $y_{l+1,k}=0$ for option k and $y_{l+1,m}=1$ for all $m\neq k$), and option $k=0$ corresponds to reporter $l+1$ remaining idle (so that $y_{l+1,m}=1$ for all m). The time complexity of this dynamic programming algorithm, called MinCost-DP, is O($NR|\mathcal{W}|$). MaxCred can also be solved, in a similar manner, using dynamic programming.

C. Leveraging the Structure of the Credibility Function

The solutions discussed so far do not leverage any *structure* in the problem. Given an event and reporter locations, the credibility associated with each report format is computed as a number and acts as an input to the algorithms discussed. However, there are two interesting structural properties in the problem formulation. First, for a given reporter at a given location, the credibility is higher for a format whose cost is also higher. Second, for reporters at different distances, the credibility decays as a function of distance. In this section, we ask the question: can we leverage this structure to devise efficient approximation algorithms, or optimal special-case solutions either for MaxCred or MinCost?

1) An Efficient Optimal Greedy MaxCred *Algorithm for Two Formats:* When a social swarming application only uses two report formats (say, text and video), it is possible to devise an optimal greedy MaxCred algorithm.

Assume each of the N reporters can report with one of two formats, f_1 or f_2, that reporters are indexed so that reporter i is closer to the event than reporter k, for $i<k$, and that credibility decays with distance. Furthermore, we assume that $e_1=\beta>1$ and $e_2=1$.

With these assumptions, the algorithm in the table below, denoted MaxCred-2F, finds an assignment with maximum credibility that falls within a budget B and runs in time $O(N^2)$.

The output of this algorithm is the maximum credibility assignment of formats to reporters.

We can prove that this algorithm is optimal.

Theorem 1: The above algorithm finds the solution C_{MAX} to MaxCred-2F problem with budget B.

We can analogously define a MinCost version for two formats, but omit it for brevity.

Algorithm 1 Algorithm MaxCred-2F

INPUT: $(c_{i,j})$: $i\in\{1,..,N\}$, $j\in\{1,2\}$; $(1,\beta)$; Budget B
Define $d_m\triangleq c_{m,1}-c_{m,2}$ for each $m\in\{1,\ldots,N\}$.
For $i\in\{0,\ldots,\min[\lfloor B/\beta\rfloor,N]\}$, do:

1) Define $Y\triangleq\min[N-i,\lfloor B-\beta i\rfloor]$.
2) Define the *active set* $\mathcal{A}\triangleq\{1,\ldots,i+Y\}$, being the set of $i+Y$ reporters closest to the event. (recall that reporters are ordered by distance).
3) Define $\mathcal{D}^*$ as the set of i reporters in $\mathcal{A}$ with the largest d_m values (breaking ties arbitrarily). Then choose format f_1 for all reporters $m\in\mathcal{D}^*$, choose f_2 for $m\in\mathcal{A}-\mathcal{D}^*$, and choose "idle" for all $m\notin\mathcal{A}$.
4) Define C^i_{MAX} as the total credibility of this assignment:
$$C^i_{MAX}\triangleq\sum_{m\in\mathcal{A}} c_{m,2}+\sum_{m\in\mathcal{D}^*} d_m$$

OUTPUT: $i^*\triangleq\arg\max_i C^i_{MAX}$.

2) A Computationally-Efficient Approximation Algorithm: In this section, we describe an approximation algorithm for MinCost, called MinCost-CC. The intuition behind MinCost-CC is that it is beneficial for a reporter to use a format that gives the highest *c*redibility per unit *c*ost (hence MinCost-CC)—this gives the most "bang for the buck." Of course, this pre-determination can result in a non-optimal choice, which is why MinCost-CC is an approximation algorithm. Formally, in MinCost-CC, reporter i chooses format f_{k^*}, where k^* is:

$$k^* = \arg\max_k\,[c_{i,k}(S_i,L)/e_k]$$

This choice can be pre-computed (since it depends only upon the credibility and cost models) with time complexity $O(NR)$, but each reporter needs to readjust its choice of the report format whenever its relative distance to the concerning event changes. The event locations that determine the format f_{k^*} chosen by a particular reporter i form annular regions about the reporter.

Once each reporter has made the format choice, it remains for the director to decide which reporter(s) to select. For MinCost-CC, the minimum cost formulation is identical to (4), and with comparable complexity, but with two crucial differences: both the constant $|\mathcal{W}|$ and the runtime now relate only to the number N of reporters, not to $N\times R$. As we shall show below, this makes a significant practical difference in runtime, even for moderate-sized inputs.

In MinCost-CC, the dynamic programming process of (5) is replaced by

$$A(l+1,s) = \max\{A(l,s), e_l + A(l,s-c_l)\} \qquad (6)$$

where c_l replaces $c_{l,j}$ in (5), since each reporter precomputes its format of choice. Compared with (5), the time complexity of (6) is reduced to $O(N|\mathcal{W}|)$ with a much smaller $|\mathcal{W}|$ in general. Notice that this time complexity is independent of R, the number of report formats, greatly improving its computational efficiency at the expense of some optimality. In addition, the overall runtime with both the time for the precomputation and the time for the dynamic programming is $O(N(R+|\mathcal{W}|))$.

Using steps similar to that presented in Section II-B, it is possible to define a MaxCred-CC approximation algorithm for maximizing credibility. We omit the details for brevity, but indicate that MinCost-CC and MaxCred-CC still have weakly-polynomial asymptotic complexity, but are computationally much more efficient than MinCost-DP and MaxCred-DP.

Evaluation of MinCost-CC. We have compared MinCost-CC with MinCost-DP in order to quantify the trade-off, for practical swarm configurations, between optimality and reduced computational complexity. Our comparison has used two datasets, one derived by manually extracting several hundred events from Google News, and a random event placement dataset generated synthetically.

We are interested in two metrics: the *optimality gap*, which is the ratio of the min-cost obtained by MinCost-CC to that obtained by MinCost-DP; and the *runtime* of the computation for each of these algorithms.

For the Google News dataset, the optimality gap is, on average 19.7%, while for random topologies, it is on average 19.0%. This is encouraging, since it suggests that MinCost-CC produces results that are not significantly far from the optimal. More interestingly, the runtime of MinCost-CC is *2-3 orders of magnitude* lower than that of MinCost-DP with the discretization setting $|\mathcal{W}| = 1000W$. This difference is not just a matter of degree, but may make the difference between a useful application and one that is not useful: MinCost-DP can take several *tens of seconds* to complete while MinCost-CC takes at most a few hundred milliseconds, which might make the difference between victory and defeat in a balloon hunt, or life and death in a disaster response swarm!

III. The Renewals Problem: Randomly Arriving Events

In the previous section we discussed a one-shot problem: that of optimizing for a single event. We now consider a sequence of events with arrival times $\{t_1, t_2, t_3, \ldots\}$, where t_k is the arrival time for event k. In this setting, we consider a stochastic variant of MaxCred, called MaxCred-Stochastic: Instead of maximizing credibility for a single event subject to a cost constraint, we maximize the average credibility-per-event subject to an average cost constraint and a per-event credibility minimum. This couples the decisions needed for each event. However, we first show that this time average problem can be solved by a reduction to individual knapsack problems of the type described in previous sections. We then show that if the per-event credibility minimum is removed, then *decisions can be made in a decentralized fashion*. The solution technique, described below, is general and can also be used to solve stochastic variants of MinCost.

A. The General Stochastic Problem

Let $\omega[k]$ represent a random vector of parameters associated with each event k, such as the location of the event and the corresponding costs and credibilities. While $\omega[k]$ can include different parameters for different types of problems, we shall soon use $\omega[k] \triangleq [(c_{i,j}[k]), (e_j[k])]$, where $(c_{i,j}[k])$ is the matrix of event-k credibility values for reporters $i \in \{1, \ldots, N\}$ and formats $f_j \in \{f_1, \ldots, f_R\}$, and $(e_j[k])$ is a vector of cost information. We assume the process $\omega[k]$ is *ergodic* with a well defined steady-state distribution. The simplest example is when $\omega[k]$ is independent and identically distributed (i.i.d.) over events $k \in \{1, 2, 3, \ldots\}$.

Let frame k denote the period of time $[t_k, t_{k+1})$ which starts with the arrival of event k and ends just before the next event. For every frame k, the director observes $\omega[k]$ and chooses a *control action* $\alpha[k]$ from a general set of feasible actions $\mathcal{A}_{\omega[k]}$ that possibly depend on $\omega[k]$. The values $\omega[k]$ and $\alpha[k]$ together determine an $M+1$ dimensional vector $\boldsymbol{y}[k]$, representing *network attributes* for event k:

$$\boldsymbol{y}[k] = (y_0[k], y_1[k], \ldots, y_M[k])$$

Specifically, each $y_m[k]$ attribute is given by a general function of $\alpha[k]$ and $\omega[k]$:

$$y_m[k] = \hat{y}_m(\alpha[k], \omega[k]) \;\; \forall m \in \{0, 1, \ldots, M\}$$

The functions $\hat{y}_m(\alpha[k], \omega[k])$ are arbitrary and are only assumed to be bounded. Define $\overline{y}_m$ as the time average expectation of the attribute $y_m[k]$, averaged over all frames:

$$\overline{y}_m \triangleq \lim_{k\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\{y_m[k]\}$$

The general problem is to find an algorithm for choosing control actions $\alpha[k]$ for each frame $k \in \{1, 2, 3, \ldots\}$ to solve:

$$\text{Minimize:} \quad \overline{y}_0 \tag{7}$$

$$\text{Subject to:} \quad 1) \;\; \overline{y}_m \leq 0 \;\; \forall m \in \{1, 2, \ldots, M\} \tag{8}$$

$$2) \;\; \alpha[k] \in \mathcal{A}_{\omega[k]} \;\; \forall \text{ frames } k \in \{1, 2, \ldots\} \tag{9}$$

The solution to the general problem is given in terms of a positive parameter V, which affects a peformance tradeoff. Specifically, for each of the M time average inequality constraints $\overline{y}_m \leq 0$ (for $m \in \{1, \ldots, M\}$) define a *virtual queue* $Z_m[k]$ with $Z_m[0] = 0$, and with frame-update equation:

$$Z_m[k+1] = \max[Z_m[k] + y_m[k], 0] \tag{10}$$

Then every frame k, observe the value of $\omega[k]$ and perform the following actions:

- Choose $\alpha[k] \in \mathcal{A}_{\omega[k]}$ to minimize:
$$V\hat{y}_0(\alpha[k], \omega[k]) + \sum_{m=1}^{M} Z_m[k]\hat{y}_m(\alpha[k], \omega[k])$$
- Update the virtual queues $Z_m[k]$ according to (10), using the values $y_m[k] = \hat{y}_m(\alpha[k], \omega[k])$ determined from the above minimization.

Assuming the problem is feasible (so that it is possible to meet the time average inequality constraints), this algorithm will *also* meet all of these constraints, and will achieve a time average value $\overline{y}_0$ that is within $O(1/V)$ of the optimum. Typically, the V parameter also affects the average size of the virtual queues (these can be shown to be $O(V)$, which directly affects the *convergence time* needed for the time averages to be close to their limiting values). The proofs of these claims follow the theory developed in [9], [10].

B. Corroboration Pull as a Stochastic Optimization Problem

Here we formulate MaxCred-Stochastic. Define $\omega[k] \triangleq [(c_{i,j}[k]), (e_j[k])]$, $\alpha[k] \triangleq (x_{i,j}[k])$, where $x_{i,j}[k]$ is a binary variable that is 1 if reporter $i \in \{1, \ldots, N\}$ uses format $f_j \in \{f_1, \ldots, f_R\}$ on frame k. The goal is to maximize the average credibility-per-frame subject to average cost constraints and to a minimum credibility level required on each frame $k \in \{1, 2, \ldots\}$:

$$\text{Maximize:} \quad \overline{c} \tag{11}$$

$$\text{Subject to:} \quad \overline{e} \leq e_{av} \tag{12}$$

$$\sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,j}[k] c_{i,j}[k] \geq c_{min} \;\; \forall \text{frames } k \tag{13}$$

$$x_{i,j}[k] \in \{0, 1\} \;,\;\; \sum_{b=1}^{R} x_{i,b}[k] \leq 1 \;\; \forall i, j, \forall \text{frames } k \tag{14}$$

where e_{av} and c_{min} are given constants, and $\overline{c}$ and $\overline{e}$ are defined:

$$\overline{c} \triangleq \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{R} \mathbb{E}\left\{x_{i,j}[k]c_{i,j}[k]\right\}$$

$$\overline{e} \triangleq \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{R} \mathbb{E}\left\{x_{i,j}[k]e_{j}[k]\right\}$$

This problem fits the general stochastic optimization framework of the previous subsection by defining $y_0[k], y_1[k]$ by:

$$y_0[k] = \hat{y}_0(\alpha[k], \omega[k]) \triangleq -\sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,j}[k]c_{i,j}[k]$$

$$y_1[k] = \hat{y}_1(\alpha[k], \omega[k]) \triangleq -e_{av} + \sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,j}[k]e_j[k]$$

and by defining $\mathcal{A}_{\omega[k]}$ as the set of all $(x_{i,j}[k])$ matrices that satisfy the constraints (13)-(14). The resulting stochastic algorithm thus defines a virtual queue $Z_1[k]$ with update:

$$Z_1[k+1] = \max\left[Z_1[k] - e_{av} + \sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,j}[k]e_j[k], 0\right] \quad (15)$$

It then observes $Z_1[k]$ and the $\omega[k]$ parameters every frame k and chooses $(x_{i,j}[k])$ subject to (13)-(14) to minimize:

$$\sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,j}[k][Z_1[k]e_j[k] - Vc_{i,j}[k]] \quad (16)$$

Algorithm performance degrades gracefully if approximate solutions to the above minimization are used [9] [10]. A simple and exact distributed implementation arises if the c_{min} constraint (13) is removed (i.e., if $c_{min} \triangleq 0$). In this case the frame k decisions are *separable over reporters* and reduce to having each reporter i choose the single format $f_j \in \{f_1, \ldots, f_R\}$ with the smallest value of $Z_1[k]e_j[k] - Vc_{i,j}[k]$, breaking ties arbitrarily and choosing to be idle (with $x_{i,j}[k] = 0$ for all $j \in \{1, \ldots, R\}$) if all of the weights $Z_1[k]e_j[k] - Vc_{i,j}[k]$ are positive. The swarm director observes the outcomes of the decisions on frame k and iterates the $Z_1[k]$ update (15), passing $Z_1[k+1]$ to all reporters before the next event occurs.

IV. Related Work

We are not aware of any prior work in the wireless networking literature that has tackled information credibility assessment. However, other fields have actively explored credibility, defined as the believability of sources or information [11], [13], [14]. Credibility has been investigated in a number of fields including information science, human communication, human-computer interaction (HCI), marketing, psychology and so on [15]. In general, research has focused on two threads: the factors that affect credibility, and the dynamics of information credibility.

The seminal work of Hovland et al. [12] may be the earliest attempt on exploring credibility, which discusses how the various characteristics of a source can affect a recipient's acceptance of a message, in the context of human communication. Rieh, Hilligoss and other explore important dimensions of credibility in the context of social interactions [11], [15], [16], such as trustworthiness, expertise and information validity. McKnight and Kacmar [11] study a unifying framework of credibility assessment in which three distinct levels of credibility are discussed: construct, heuristics, and interaction. Their work is in the context of assessing the credibility of websites as sources of information.

Finally, there is a body of work that has examined processes and propagation of credible information. Corroboration as a process of credibility assessment is discussed in [17]. Proximity, both geographic and social, and its role in credibility assessment is discussed in [5]: our role of geographic distance as a measure of credibility is related to this discussion. Saavedra et al. [18] explore the dynamics and the emergence of synchronicity in decision-making when traders use corroboration as a mechanism for trading decisions.

V. Conclusions and Future Work

In this paper, we have explored the design space of algorithms for a new problem, optimizing corroboration pull in an emerging application area, social swarming. We have proposed optimal special-case algorithms, computationally efficient approximations, and decentralized optimal stochastic variants. Several directions for future work are possible: increasing credibility and cost model realism, incorporating malice, allowing peers to relay reports, and exploring other realistic, yet efficient and near-optimal special-case solutions.

References

[1] M. Min, S. Reddy, K. Shilton, N. Yau, and et al. PEIR, the personal environmental impact report, *Proc. 7th ACM Mobisys*, June 2009.

[2] E. Miluzzo, N. Lane, K. Fodor, R. Peterson, and et al. Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application, *Proc. 6th ACM SenSys*, November 2008.

[3] S. Kang, J. Lee, H, Jang, H. Lee, and et al. SeeMon: scalable and energy-efficient context monitoring framework for sensor-rich mobile environments. *Proc. 6th ACM MobiSys*, June 2008.

[4] R. K. Balan, D. Gergle, M. Satyanarayanan, J. Herbsleb. Simplifying cyber foraging for mobile devices, *Proc. 5th ACM MobiSys*, June 2007.

[5] M. T. Rivera, S. B. Soderstrom, and B. Uzzi. Dyads: Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annual Review of Sociology*, Vol. 36, pp 91-115, Jun 2010.

[6] M. Ra, J. Paek, A. B. Sharma, R. Govindan, M. H. Krieger, and M. J. Neely. Energy-Delay Tradeoffs in Smartphone Applications. *Proc. 8th ACM Mobisys*, June 2010.

[7] S. Martello and P. Toth. *Knapsack Problems: Algorithms and Computer Implementations*. Wiley: Chichester, England, 1990.

[8] R.K. Ahuja, T.L. Magnanti, J.B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

[9] L. Georgiadis, M. J. Neely, and L. Tassiulas. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, vol. 1, no. 1, pp. 1-149, 2006.

[10] M. J. Neely. *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan & Claypool, 2010.

[11] B. Hilligoss and S. Rieh. Developing a Unifying Framework of Credibility Assessment: Construct, heuristics, and Interaction in Context. *Information Processing and Management*, Vol. 44, 1467-1484, 2008.

[12] C. I. Hovland and W. Weiss. The Influence of Source Credibility on Communication Effectiveness. *Public Opinion Quarterly*, Vol. 15, pp 635-650, 1951.

[13] D. H. McKnight and C. J. Kacmar. Factors and Effects of Information Credibility. *Proc. ICEC'07*, Minneapolis, 2007.

[14] B. J. Fogg and H. Tseng. The Elements of Computer Credibility. *Proc. CHI'99*, 2004.

[15] S. Rieh and D. Danielson Credibility Models for Multi-Source Fusion. *Annual Review of Information Science and Technology*, Vol. 41, pp 307-364, 2007.

[16] P. Chen. Information Credibility Assessment and Meta Data Modeling in Integrating Heterogeneous Data Sources. *Air Force Research Laboratory Technical Report AFRL-IF-RS-TR-2002-298*, 2002.

[17] R. B. Cialdini. The Science of Persuasion. *Scientific American Mind*, Jan 2004.

[18] S. Saavedra, B. Uzzi, and K. Hagerty. Synchronicity and the collective genius of profitable day traders. Working Paper, Kellogg School of Management, 2010.