institute IMdea networks

# technical report

TR-IMDEA-Networks-2013-1

# Simple Approximate Analysis of Floating Content for Context-Aware Applications

Shahzad Ali

Gianluca Rizzo

Balaji Rengarajan

Marco Ajmone Marsan

May 2013

# A Simple Approximate Analysis of Floating Content for Context-Aware Applications

Shahzad Ali
Institute IMDEA Networks, Madrid, Spain
Universidad Carlos III de Madrid, Spain
shahzad.ali@imdea.org

Gianluca Rizzo
HES-SO Valais, Sierre, Switzerland
gianluca.rizzo@hevs.ch

Balaji Rengarajan
Institute IMDEA Networks, Madrid, Spain
balaji.rengarajan@imdea.org

Marco Ajmone Marsan
Institute IMDEA Networks
Madrid, Spain
Politecnico di Torino, Italy
marco.ajmone@imdea.org

## ABSTRACT

Context-awareness is a peculiar characteristic of an ever expanding set of applications that make use of a combination of restricted spatio-temporal locality and mobile communications, to deliver a variety of services to the end user. Communication requirements for context-aware applications significantly differ from those of ordinary applications; opportunistic communications are extremely well-suited to them, because they naturally incorporate context. Recently, an opportunistic communication paradigm called "Floating Content" was proposed, which is conceived to support serverless, distributed content sharing. In this work, we present a simple (in that it uses few primitive system parameters), approximate analytical model for the performance analysis of context-aware applications that use floating content. From a system design perspective, our analysis can be used to tune key system parameters so as to achieve the desired application performance. In particular, we apply our analysis to estimate the "success probability" for two representative categories of context-aware applications, and show how the system can be configured to achieve the application's target. In order to complement our analytical study, we validate our model using extensive simulations under different settings and mobility patterns. Our simulation results show that our model-based predictions are indeed highly accurate under a wide range of conditions.

## 1. INTRODUCTION

The growth of mobile computing, and the pervasiveness of smart user devices is progressively driving applications towards context-awareness, i.e., towards applications and services that allow users to exploit "any information that can be used to characterize the situation of an entity" [1]. This enables applications to improve their efficiency and utility, and to offer services that better suit the needs of users in a given situation.

One of the best examples of context, and one that is widely used by context-aware applications is spatial and temporal locality. As an example, consider a context-aware parking finding application [2]. Information about a vacant parking spot may be of interest for a limited time (until the space is filled), and only to users who are in fairly close proximity. Similarly, a shop might want to spread advertisements about a sale only during the sale period and to customers in the neighborhood. Many more examples of context-aware applications are emerging, that make use of spatio-temporal locality and wireless communications to deliver a variety of services. By the end of 2010, there were more than $100,000$ applications developed for the iPhone alone, and about $10\%$ of them employed localization technology [2]. It is expected that by 2014 more than 1.5 billion people would be using applications based on local search (search restricted on the basis of spatio-temporal locality), and that mobile location based services will drive revenues of more than \$15 billion worldwide [3].

A common feature of context-aware applications is that they have communication requirements that significantly differ from ordinary applications. For most location-based, context-aware applications, the scope of generated content itself is local. This locally relevant content may be of little concern to the rest of the world, therefore moving this content from the user device to store it in a well-accessible centralized location and/or making this information available beyond its scope represents a clear waste of resources (connectivity, storage). Due to these specific requirements, opportunistic communication can play a special role when coupled with context-awareness. The benefit of opportunistic communications is that it naturally incorporates context as spatial proximity is closely associated with connectivity.

In our work we consider a specific opportunistic communication paradigm, known as *floating content* (FC) [4], conceived to support server-less distributed content sharing. It aims at ensuring the availability of data within a certain geographic area called *anchor zone* (AZ), and for a given duration in time. Within the AZ, any time a user who is unaware of the content enters the transmission range of another user possessing it, the content is shared. Thus, the content can be replicated on a set of nodes within the AZ, who in turn will pass it on to other nodes that come into their range before leaving the AZ. As a result, information might be stored on some nodes within the AZ even after the original node which generated it has left, i.e., content 'floats' within the AZ. Users who traverse the AZ while content is floating have an opportunity to learn the content, provided they meet a node with information prior to leaving the AZ.

Due to stochastic fluctuations of the number of nodes with content within the AZ, and to their traversal patterns, the content ultimately disappears from the AZ. The duration of time for which content stays available in the AZ (at least one node in the AZ has the content) is the floating lifetime. While in [4] and [5], the focus is on understanding the asymptotic properties of the floating lifetime, our objective is to characterize the performance of context-aware applications using the FC paradigm. We restrict ourselves to the regime where floating lifetime is expected to be large and study the *success probability*, a key performance indicator that captures the likelihood of intended users to receive the relevant information. A key characteristic of our modeling approach is that success probability is computed from few primitive system parameters, most notably the probability density function of the length of the path followed by users within the AZ. This allows the analysis to be generalized to settings different from the one we consider in this paper, including different AZ shapes, different user mobility patterns, different user speed distributions, different service and application models. Our main contributions are:

- We develop an approximate analytical model for success probability, with key parameters the AZ radius, the node density, and the node transmission range.
- We apply our model to two representative categories of context-aware applications, and derive expressions for their success probability.
- We demonstrate how the model predictions can be used to tune key system parameters to achieve the desired application performance
- We validate our analysis with extensive simulations using OMNeT++ [6]. Simulation results reveal that the predicted success probability is indeed very accurate.

The rest of the paper is organized as follows. We briefly discuss related work in Section 2, emphasizing differences with respect to our work. In Section 3, we present the system model for FC and define the application performance measures. In Section 4, we present our approximate analysis, and in Section 5, we validate our model via simulations. Finally, we present our conclusions in Section 6. In order to improve the QoE of the reader, we omit proofs in the body of the paper, including them in appendices.

## 2. RELATED WORK

In [4], the authors introduced the concept of Floating Content, and derived asymptotic conditions (called criticality conditions) for the expected floating lifetime to be large under some large population assumptions. This corresponds to conditions for the information to remain available in the AZ, and supports the viability of the FC paradigm. The criticality condition depends on three key parameters: the average number of nodes in the AZ, the average contact rate experienced by a node, and the average sojourn time of a node within an AZ. In [5], the authors validated the analytical results presented in [4] with extensive simulations, and showed that FC is feasible even when there are modest number of nodes in the network. An open issue in these papers is the lack of a correlation between the primary performance parameters from an application perspective and the main design parameters of FC. For concrete applications, it is not sufficient that the content asymptotically floats: for the application performance are vital the density of nodes with content inside the AZ, their spatial distribution, and the percentage of times a node gets the content, once it enters the AZ. Therefore, in this paper we address the problem from a different perspective with respect to [4] and [5], and investigate the effect of the system design parameters on the performance of an application using FC.

The concept of Floating Content is not new, and in recent years concepts similar to it have appeared in the literature with different names. A system called *Locus* is proposed in [7]. In [8], a similar concept called *Hovering Information* is presented, and two algorithms to improve information availability are presented. Both Locus and Hovering Information are quite simila to Floating Content. [9] investigates the benefit accrued from limited infrastructure support. In [10], the publish-subscribe paradigm is used to build an opportunistic spatio-temporal dissemination system for vehicular networks, while the amount of time for which information stays available in a vehicular network was investigated in [2]. However, unlike this work, the performance from a user perspective is not the focus of the aforementioned papers.

# 3. SYSTEM MODEL

We consider mobile nodes in $\mathbb{R}^2$, that form a homogeneous Poisson process with intensity $\lambda$ nodes per square meter at time $t = 0$.

**Mobility Model:** We assume that node movements follow the Random Direction (RD) mobility model [11], in which nodes independently travel along a straight line, with an angle of movement uniformly distributed between 0 and $2\pi$. We assume that nodes move with a constant velocity $v$ m/s. In [12], it is shown that under this mobility model, the spatial node distribution remains uniform at all time instants. The reason for choosing this mobility model is its simplicity and analytical tractability. We explore the impact of more realistic mobility through simulations in order to determine if our simple model is adequate to capture the first-order effects.

**Content Replication:** At any time instant, a node with an information item $\mathcal{I}$ can seed an *anchor zone* (AZ), which is a circular area with radius $R$ meters centred at its current position. Initially, the seeder is the only node with the content $\mathcal{I}$. Every time two nodes *within the AZ* come in transmission range of each other (we call this a contact), if only one of the two nodes possesses $\mathcal{I}$, it communicates $\mathcal{I}$ to the other. We assume there is no supporting infrastructure available, so that nodes must rely exclusively on ad-hoc communication. All nodes are assumed to have the same transmission range of $r$ meters. We assume that $R >> r$, since these are the cases where content floating has practical utility. The probability that information is transferred during a contact is denoted by $Q$, and takes into account transmission errors, collisions, and contact duration. Note that nodes delete their own copy of $\mathcal{I}$ when they move outside of the AZ.

**Performance Metric:** In this paper, we focus on the scenarios where the node density and the anchor zone radius are large with respect to the transmission radius, resulting in large floating lifetimes on average. In particular, we assume that the criticality condition derived in [4] holds, so that the expected lifetime of content floating is infinite under the fluid limit approximation of [4]. The measure of performance that we use is the probability that a node entering an anchor zone within the floating lifetime, i.e., when at least one node within the anchor zone has content, successfully receives the information item $\mathcal{I}$.

Definition 1 (Success probability). *The success probability $P_s(\tau)$ is the probability that a node receives the content $\mathcal{I}$ associated with an AZ within a time $\tau$ after entering the AZ (if it is still in the AZ), or by the time it leaves the AZ (if it leaves it before time $\tau$) conditional on there being at least one node within the AZ with content at the time of the node's entry.*

Note that $P_s(\infty)$ is the probability that a node entering the anchor zone within the floating lifetime receives the content before leaving, and will be denoted as $P_s$. We empirically measure this quantity by simulating a number of AZs and tracking the fraction of nodes that enter within the floating lifetime, and successfully receive the content.

# 4. ANALYSIS

As a first step, we compute an expression for approximating success probability for the general floating content model described in Section 3. Then, we extend our analysis to include two different categories of applications, and derive approximate expressions for success probability for these applications.

**General Floating Content:** We assume in our derivation that the system is in an equilibrium state (as assumed in [4]), where the average rate of nodes without content $\mathcal{I}$, which get $\mathcal{I}$ inside the AZ, is equal to the average rate at which nodes having $\mathcal{I}$ leave the AZ. In such an equilibrium, the average number of nodes with and without $\mathcal{I}$ within the AZ remains constant. Note that this equilibrium assumption will be a first source of discrepancy with respect to simulations, where we observe periods in which the numbers of nodes with and without content fluctuate (e.g., the number of nodes with content typically grows right after the appearance of a seeder node at the center of the AZ).

Result 1. *Consider an AZ with radius $R$, node density $\lambda$, and nodes with transmission range $r$ and speed $v$. Let $Q$ denote the probability that two nodes successfully transfer the content $\mathcal{I}$ while they are in contact. Then $P_s(\tau)$ for $\tau \le 2R/v$ can be approximated as*

$$P_s(\tau) = \int_0^{2R} \frac{l^2}{\pi R^2 \sqrt{4R^2 - l^2}} \cdot$$

$$\cdot \sum_{k=1}^{\infty} \left[ 1 - \left( 1 - \frac{Q\overline{n}}{(\overline{m} + \overline{n})} \right)^k \right] \frac{(2r\lambda(l \wedge v\tau))^k e^{-2r\lambda(l\wedge v\tau)}}{k!} dl \tag{1}$$

*where $\overline{m} = \min(\frac{v}{Q\nu R}, \lambda\pi R^2)$, $\overline{n} = \lambda\pi R^2 - \overline{m}$, with $\nu$ given by $\frac{2rv^2}{(\pi R^2)}$.*

Here, $a \wedge b$ stands for $\min(a, b)$. $\overline{n}$ and $\overline{m}$ are respectively the average number of nodes with and without content within the anchor zone. The derivation of this result is presented in the Appendix.

Note that in the expression above, the integral is over $l$ which is the length of the AZ chord traversed by a node. The expression calculates the probability that a node meets $k$ other nodes during its traversal as the product of the pdf of the chord length and the conditional pdf of the number of contacts given the chord length. The term in square brackets is the probability that at least one out of the $k$ nodes met has the content, and that the transfer is successful. In deriving the above result, we assume that the distribution of nodes with content in the AZ is uniform, and that the odds of meeting nodes possessing the information are uncorrelated. In reality, both assumptions are not satisfied, since there is some spatial clustering of nodes with content, and also a higher density of such nodes near the center of the AZ. However, as we show through simulations in the sequel, these are second-order effects and the result in (1) captures the success probability well.

For some applications of floating content, more important than computing the probability that a node obtains content $\mathcal{I}$ within a time $\tau$ after entering the AZ, is estimating the probability that a node obtains the content before leaving the AZ. We denote this probability as $P_s$.

COROLLARY 1. *With the same assumptions as in Result 1, the success probability $P_s$ can be approximated as $P_s(2R/v)$.*

In the derivation of Result 1, this corresponds to integrating over the whole length of the chord travelled within the AZ.

In these results, all factors influencing the probability of successful content transfer between two nodes in range of each other are captured by the parameter Q. This allows detailed models for information transfer at physical and/or MAC layer to be easily incorporated in the analysis, without changing the structure of the formula. In the following theorem we provide a simple expression for the probability of successful content transfer, which accounts for finite bandwidth availability and transmission errors, assuming that nodes continually retry on failure as long as they stay within transmission range.

THEOREM 1. *The probability of successful ransfer of content between two nodes, assuming that the minimum required time for the transfer is $X'$, is given by*

$$Q(X') = \sum_{k=1}^{\infty} \int_{kX'}^{(k+1)X'} [1-(1-S)^k] f_\tau(t)dt \qquad (2)$$

*where $f_\tau(t)$ is the probability density function of the contact duration under the RD mobility model, given by*

$$f_\tau(t) = \int_0^{min(2v, \frac{2r}{t})} \frac{2\omega^3 t^2}{\pi^2 r^2 \sqrt{4r^2 - \omega^2 t^2}\sqrt{4v^2-\omega^2}} d\omega \qquad (3)$$

*and $S$ is the probability of no transmission failures (errors, collisions, etc) for each content transfer attempt.*

The derivation of this result is presented in the Appendix.

**Application-specific analysis:** The performance parameter "success probability" which we considered so far is relevant when FC is used to ensure that users traversing a given AZ (we assumed it to be circular, but extensions to different shapes are simple) get the associated content. However, in order to ensure acceptable application performance, content could be floated in a geographic area that is a superset of the area where it is needed. Therefore, we consider an AZ with radius $R_2$ which acts as the "replication range" within which content is replicated using FC. A new zone with radius $R_1$, with $R_1 \leq R_2$, called Range Of Interest (ROI) is also defined. ROI is application-specific and depends on the particular service which has to be delivered. Note that the absolute and relative values of $R_1$ and $R_2$ can be chosen so as to achieve the desired system (rather, application) performance. Below, we consider two application categories that have different interpretations for the successful delivery of content.

**Application category 1:** For some applications, it is important to deliver a message to end users when they get within a given range, because the message is expected to trigger some specific actions, like visiting a famous tourist attraction, or a restaurant. One example of such application can be advertising, when it is desired that a user should be notified about some offer/discount before leaving a certain geographic area. The expression for success probability for this application category is calculated in Result 2 using a new definition for success probability.

DEFINITION 2. *The success probability for getting content $\mathcal{I}$ before leaving the ROI ($P_{SBL}$) is the probability that a node entering the ROI gets the content $\mathcal{I}$ before exiting the ROI, conditional on the presence of at least a single node with content in the AZ at the time of the node's entry.*

RESULT 2. *For an AZ with radius $R_2$ and a ROI with radius $R_1$, with $R_1 \leq R_2$, $P_{SBL}$ can be approximated as*

$$P_{SBL} = \sum_{k=1}^{\infty} \left( \int_{\sqrt{R_2^2 - R_1^2}}^{R_1+R_2} \left( f_{L_1}(\ell) \frac{(2r\ell\lambda)^k e^{-2r\ell\lambda}}{k!} \right) d\ell \right) \left[ 1 - \left( 1 - \frac{Q\overline{n}}{(\overline{m}+\overline{n})} \right)^k \right] \qquad (4)$$

*where $f_{L_1}(l)$ is given by*

$$f_{L_1}(l) = \frac{2\sqrt{R_2^2 - (g^{-1}(l))^2}\sqrt{R_2^2 - (g^{-1}(l))}}{(\int_0^{R_1} g(y)dy) g^{-1}(l)} \qquad (5)$$

*and* $g(y) = \sqrt{R_2^2 - y^2} + \sqrt{R_1^2 - y^2}$.

The proof of Result 2 is presented in the Appendix.

**Application category 2:** For a different type of application, it is important to deliver the content $\mathcal{I}$ to users before they enter a certain area (which we identify again with the ROI). Examples of such applications can be accident or traffic jam warnings, when a user should be notified about the situation before entering a certain geographic area so that he/she can make an informed decision about alternate paths. The expression for success probability for such kind of applications is derived in Result 3, using yet another definition for success probability.

DEFINITION 3. *The success probability for getting content $\mathcal{I}$ before entering the ROI ($P_{SBE}$) is the probability that a node entering the AZ gets the content $\mathcal{I}$ before entering the ROI, conditional on the presence of at least a single node with content in the AZ at the time of the node's entry.*

RESULT 3. *For an AZ with radius $R_2$, and a ROI with radius $R_1$, with $R_1 \leq R_2$, $P_{SBE}$ can be approximated as*

$$P_{SBE} = \sum_{k=1}^{\infty} \int_{R_2-R_1}^{\sqrt{R_2^2-R_1^2}} \left( f_{L_2}(\ell) \frac{(2r\ell\lambda)^k e^{-2r\ell\lambda}}{k!} \right) d\ell \left[ 1 - \left( 1 - \frac{Q\overline{n}}{(\overline{m}+\overline{n})} \right)^k \right] \quad (6)$$

*where $f_{L_2}(l)$ is given by*

$$f_{L_2}(l) = \frac{2\sqrt{R_2^2 - (h^{-1}(l))^2}\sqrt{R_2^2 - (h^{-1}(l))}}{(\int_0^{R_1} h(y)dy)h^{-1}(l)} \quad (7)$$

*and* $h(y) = \sqrt{R_2^2 - y^2} - \sqrt{R_1^2 - y^2}$.

The proof of Result 3 is presented in the Appendix.

## 5. SIMULATIONS AND RESULTS

In this section, we present and discuss numerical results obtained from our approximate models of the previous sections, and from simulations. For all simulation experiments, the OMNeT++ based framework called INET [6] is used. Confidence intervals at 95% confidence level were evaluated for all cases using independent replications, and are shown in the following figures, with the exception of Fig. 2, in order to avoid cluttering the graph. The purpose of the presentation of numerical results is twofold. On the one hand, we validate the approximate expressions derived for success probability, showing that they are accurate under varied conditions. On the other hand, we show the effectiveness of the analysis in selecting the Floating Content parameters for different applications and under different scenarios. In simulations, the average user density ranges from 22 to 66 nodes per square kilometer. The transmission range is 50 meters and nodes move with a constant speed of 10 meters per second. Anchor zone radius ranges from 500 to 1000 meters. We simulate multiple instances of anchor zones, and measure success probability in each instance over the floating lifetime or until a maximum of 50000 seconds have elapsed.

**Results for General Floating Content:** Fig. 1 shows both the analytical predictions and the empirically determined values of success probability as a function of the AZ radius. In addition to the random direction mobility model (RDMM), we evaluate success probability under the generalized Manhattan mobility model (MGMM) with block sizes of 100m $\times$ 150m, and probability of turning left or right at an intersection set to be 0.25 each and that of heading straight set to 0.5. To model success probability for MGMM, we empirically measure the average number of nodes met by a node while traversing the AZ as well as the overall rate of contacts. We assume, as we did in the analysis of the RDMM, that node contacts are uniformly distributed in space. In order to compute the predicted success probability, we use an analog of Result 1 with the number of contacts during a node traversal modeled by a Poisson random variable parametrized with the empirical mean.

It can be seen that the model predictions match very well with the simulated results for both RDMM and MGMM, suggesting that the model indeed captures successfully the first order-effects on success probability. The curves in the figure also show that an increase in both node density and AZ radius improve the success probability. An increase in the the AZ radius increases the average time a node spends inside the AZ, therefore the chances of meeting a node having content $\mathcal{I}$ increase. Similarly, an increase in node density also results in higher overall contact rate as well as a higher chance of meeting a node with content $\mathcal{I}$. For identical node density (33 nodes per square kilometer), under MGMM the success probability is higher than RDMM. The major reason behind this is the higher contact rate under MGMM compared to RDMM. This increases the population of nodes with content in the AZ, and also the odds of a node meeting a node with content under MGMM. Fig. 1 also shows the impact of transmission errors and finite bandwidth for RDMM. It can be seen that under a finite bandwidth model with a data rate of 11 Mbps, for transmission error probability of 0.2 with a file size of 2MB, the success probability decreases. As the transmit errors and limited contact times reduce the rate of contacts where communication is successful, this reduces the fraction of nodes in the AZ with content as well as the overall success probability.

Fig. 2 shows success probability versus the node den-

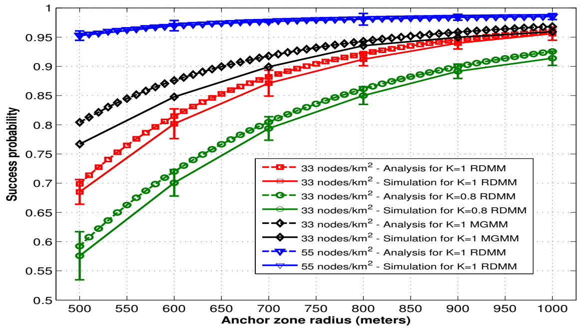


**Figure 1: Success probability vs. AZ radius.**

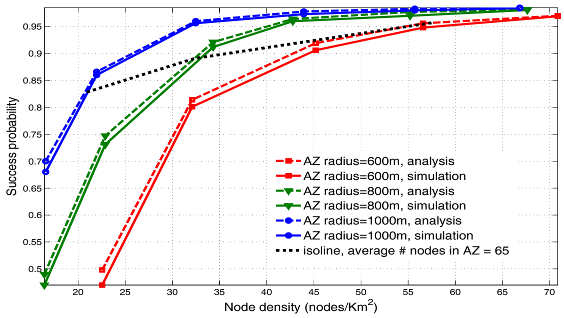


**Figure 2: Success probability vs. node density.**

sity for different choices of AZ radius. As the node density increases, a large improvement in success probability can be observed. The analytical model captures this effect, and is a very good predictor for success probability. Further, Fig. 2 shows that as AZs grow larger, a given success probability threshold can be achieved at lower node densities. This is due to node paths through the AZ getting larger with AZ size, resulting in more opportunities to obtain content. However, if node densities and AZ size are varied jointly such that the average number of nodes in the AZ stays unchanged (see the isoline corresponding to an average of 65 nodes in the AZ), larger AZs result in lower success probabilities. The key parameter behind this effect is the ratio between the AZ radius ($R$) and the node transmission range ($r$). Indeed, as the AZ radius increases, this ratio decreases, and as a result more nodes are needed in the AZ in order to achieve a similar success probability. Clearly, defining very large AZ might lead to wastage of resources without significantly improving performance. Thus, the ratio of the AZ radius to the transmission range is a critical parameter, that must be tuned carefully.

**Results for Applications:** Figs. 3 and 4 show curves of success probability versus the AZ radius $R_2$, for different node densities, for application categories 1 and 2, when ROI radius ($R_1$) is 200m. Increases in either AZ radius ($R_2$), or node density result in increased success probability for both applications. In both cases, nodes traversing the AZ make more contacts resulting in a larger fraction of nodes having content in the AZ,

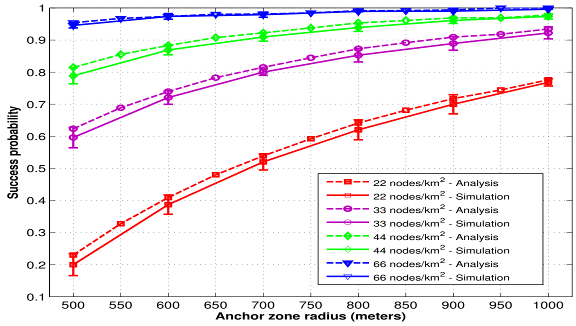


**Figure 3: Success probability for application 1 with ROI = 200m.**

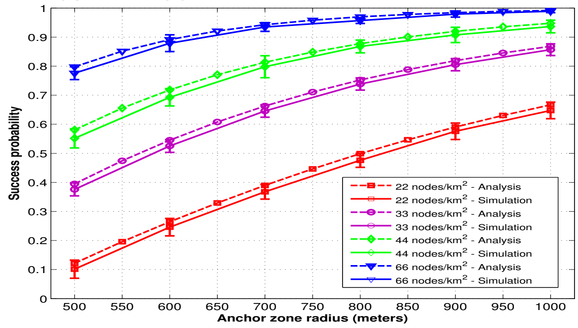


**Figure 4: Success probability for application 2 with ROI = 200m.**

and more opportunities for a node to obtain the content. We see that increasing AZ radius has diminishing returns, with this effect being more pronounced in the case of application 1. It can be seen from both figures that our analytical results are very close to simulations results.

In a real-world environment, an application using FC is likely to require a minimum success probability (a shop may require that a given percentage of the people passing within 200 m of its premises receive their advertisement). However, unnecessarily large AZs would lead to resource wastage. From Figs. 3 and 4, we see that the proposed analytical model can be used in order to tune AZ radius to achieve the desired success probability. Here, we assume the threshold to be 90%, and use our models to compute the minimum AZ radius ($R_2$) that is required to achieve this objective at different ROIs and node densities. Figs. 5 and 6 depict the results.

For application 1, in Fig. 5 we can observe that, for a given node density, as ROI increases, the required $R_2$ decreases until the condition is reached where replicating the content within the ROI is sufficient to achieve the desired success probability. When such condition is reached, $R_1$ becomes equal to $R_2$ (we always consider $R_1 \leq R_2$). In the case of application 2, we observe from Fig. 6 that, for a given density of nodes, the required AZ size ($R_2$) increases as $R_1$ increases. Here, as the ROI increases, we also need larger $R_1$ in order to make sure

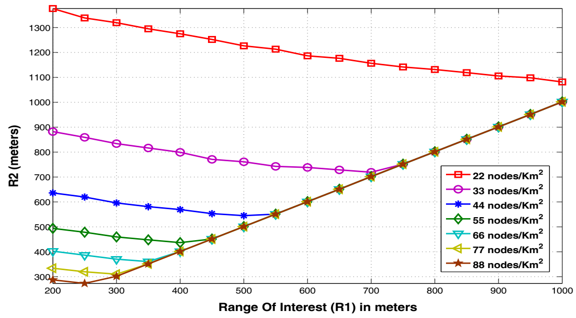


**Figure 5: AZ radius to achieve 90% success probability for application 1**

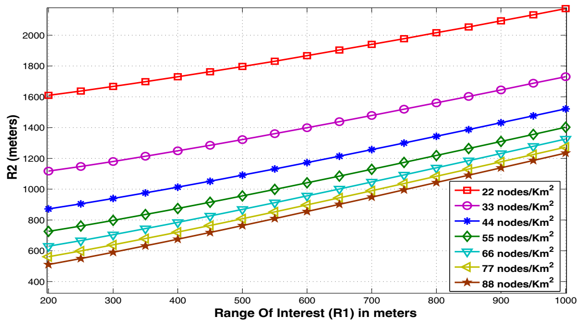


**Figure 6: AZ radius to achieve 90% success probability for application 2.**

that node paths in the AZ are long enough for them to learn the content with sufficiently high probability before entering the ROI.

## 6. CONCLUSIONS

In this paper, we focus on floating content and its ability to act as the communication paradigm supporting context-aware applications. We defined success probability as the primary performance indicator, and developed a simple, approximate analytical modeling framework, which can be adapted to several different settings. Indeed, our models essentially depend on the length of the path within the AZ where content floats and the spatial density of nodes. Different settings (in terms of AZ shape, user speed distribution, application or service characteristics, etc.) may correspond to different path lengths, but the analytical approach we presented in this paper remains applicable. Both for the Random Direction and the Manhattan mobility models, our approximation computes very accurate success probability values for a wide range of anchor zone radii and node densities, as proved by comparison against the results of detailed simulation experiments.

Our models can be adapted to several different categories of context-aware applications, and the model predictions can be used in order to tune key parameters of the system to achieve the required performance with minimum overhead. We studied two such cases in this paper, deriving approximate expressions for success probability for both. In addition, in these two cases, we validated our model using extensive simulations in OMNeT++, proving the very good accuracy of our analytical predictions. Our simulation results show that high success probabilities are achieved with reasonably sized anchor zones that are only slightly larger than the region of interest even at low node densities, demonstrating the viability of floating content as an enabler for context-aware applications.

## 7. REFERENCES


[1] A. Zimmermann, A. Lorenz, and R. Oppermann, "An operational definition of context," in *CONTEXT'07*, 2007, pp. 558–571.

[2] B. Liu, B. Khorashadi, D. Ghosal, C.-N. Chuah, and M. H. Zhang, "Assessing the vanet's local information storage capability under different traffic mobility," in *INFOCOM'10*, 2010, pp. 116–120.

[3] A. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: challenges faced in developing techniques for accurate wireless location information," *Signal Processing Magazine, IEEE*, vol. 22, no. 4, pp. 24 – 40, July 2005.

[4] E. Hyytiä, J. Virtamo, P. Lassila, J. Kangasharju, and J. Ott, "When does content float? characterizing availability of anchored information in opportunistic content sharing," in *INFOCOM*, Shanghai, China, Apr. 2011, pp. 3123–3131.

[5] J. Ott, E. Hyytia, P. Lassila, T. Vaegs, and J. Kangasharju, "Floating content: Information sharing in urban areas," in *PerCom 2011*, March 2011, pp. 136 –146.

[6] A. Varga, "The omnet++ discrete event simulation system," *ESM 2001*, June 2001.

[7] N. Thompson, R. Crepaldi, and R. Kravets, "Locus: a location-based data overlay for disruption-tolerant networks," in *CHANTS '10*, 2010, pp. 47–54.

[8] A. A. V. Castro, G. Di Marzo Serugendo, and D. Konstantas, "Hovering information - self-organising information that finds its own storage," in *SUTC '08*, 2008, pp. 193–200.

[9] J. Whitbeck, Y. Lopez, J. Leguay, V. Conan, and M. D. de Amorim, "Push-and-track: Saving infrastructure bandwidth through opportunistic forwarding," *Pervasive and Mobile Computing*, 2012.

[10] I. Leontiadis and C. Mascolo, "Opportunistic spatio-temporal dissemination system for vehicular networks," in *MobiOpp '07*, 2007, pp. 39–46.

[11] C. Bettstetter, "Mobility modeling in wireless networks: Categorization, smooth movement, and

border effects," *ACM Mobile Computing and Communications Review*, p. 2001.

[12] P. Nain, D. Towsley, B. Liu, and Z. Liu, "Properties of random direction models," in *INFOCOM '05*, vol. 3, March 2005, pp. 1897 – 1907.

[13] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Performance analysis of mobility-assisted routing," in *MobiHoc '06*, 2006, pp. 49–60.

[14] M. D. Springer, *The Algebra of Random Variables.* John Wiley & Sons, 1979.

## APPENDIX

**Proof of Result 1:** In order to prove Result 1, we first introduce the following lemmas.

LEMMA 1. *Under the RD mobility model with node density $\lambda$, when nodes have a transmission radius of $r$, and velocity equal to $v$, the number of contacts made by a node in a time interval $\tau$ is Poisson distributed with mean $\mu_C = 2rv\tau\lambda$.*

PROOF. Without loss of generality, we consider the perspective of a node $i$ with velocity vector $\vec{v}_i = (v, \angle 0)$. We calculate the relative velocities of all nodes in the system with node $i$ as the reference. Thus, we can consider node $i$ to be non-moving, with the other nodes mobility characterized by their relative velocity vector. The total number of new contacts made by $i$ in a time interval $\tau$ is equivalent to the number of nodes that enter a circle of radius $r$ centred at the non-moving node $i$ in that time interval, under the relative velocity model. Under the RD mobility model, nodes never change direction and thus a single node cannot enter this circle multiple times. Nodes choose their directions independently and there are no spatial correlations under the RD model, thus the number of nodes entering the circle in two disjoint time intervals are also independent. Further, the stationarity of the RD model implies that the distribution of the number of nodes entering the circle only depends on the interval $\tau$. Therefore, the number of contacts made by a node in a time interval is Poisson distributed.

In order to find the mean of the Poisson distribution, we integrate the average rate at which nodes cross the boundary of the circle delineating the transmission range. The relative velocity of node $j$ with velocity vector $\vec{v}_j = (v, \angle\theta_j)$ is

$$\vec{v'}_j = \left(2v\sin\left(\frac{\theta}{2}\right), \angle\frac{\pi+\theta}{2}\right)$$

We consider infinitesimal sections of the circumference of the circle, which lie at an angle $\psi$ with respect to the x axis. At each section, we consider the possible directions from which nodes could enter (note that relative velocities vectors have angles between $\pi/2$ and $3\pi/2$), and use the relative velocity to capture the flow from each direction entering the transmission range through that section, The mean number of contacts in the interval $\tau$ is:

$$\int_0^{\tau}\int_0^{\pi}\int_{\psi+\frac{\pi}{2}}^{\frac{3\pi}{2}} 2v\sin\left(\frac{2\phi-\pi}{2}\right)\cos(\phi-\psi)\frac{\lambda r}{\pi}d\phi d\psi dt$$

$$+\int_0^{\tau}\int_{\pi}^{2\pi}\int_{\frac{\pi}{2}}^{\psi-\frac{\pi}{2}} 2v\sin\left(\frac{2\phi-\pi}{2}\right)\cos(\phi-\psi)\frac{\lambda r}{\pi}d\phi d\psi dt$$

$$= 2rv\tau\lambda$$

□

LEMMA 2. *In equilibrium state, for a node traversing an AZ under the Random Direction mobility model, the probability of meeting $k$ nodes is given by*

$$\int_0^{2R} \frac{\ell^2}{\pi R^2\sqrt{4R^2-\ell^2}}\frac{(2r\ell\lambda)^k e^{-2r\ell\lambda}}{k!}d\ell \tag{8}$$

PROOF. Let A and B be the entry and exit points of a node traversing an AZ, respectively, as shown in Fig. 7. Since nodes move in a straight line, the length of the chord AB is given by

$$L(y) = 2\sqrt{R^2-y^2} \tag{9}$$

where $y$ is the distance |CY|. Due to the properties of the RD mobility model, node distribution is uniform in space at any point in time [12], and therefore $y$ is uniformly distributed in $[0, R]$. Then the pdf of $L$ is given by

$$f_L(y) = \frac{2\sqrt{R^2-y^2}}{\int_0^R 2\sqrt{R^2-y^2}dy} = \frac{4\sqrt{R^2-y^2}}{R^2\pi} \tag{10}$$

Letting $\ell = L(y)$, and substituting we finally get

$$f_L(\ell) = \frac{f_L(g_1^{-1}(\ell))}{|g_1'(g_1^{-1}(\ell))|} = \frac{\ell^2}{R^2\pi\sqrt{4R^2-\ell^2}} \tag{11}$$

with $\ell \in [0,\ 2R]$.

Using Lemma 1 with $l = v\tau$, the probability of meeting $k$ nodes along this trajectory with a Poisson distribution, with intensity $2r\ell\lambda$:

$$P(meet\ k\ nodes|\ell) = \frac{(2r\ell\lambda)^k e^{-2r\ell\lambda}}{k!} \tag{12}$$

So that the probability of meeting $k$ nodes is given by (8). □

LEMMA 3. *Consider an AZ in equilbrium state, with an average number of nodes equal to $\overline{N}$, and let $\overline{n}$ and $\overline{m}$ denote the average numbers of nodes with and without content $\mathcal{I}$, respectively. Then $\overline{m} = \min(\frac{v}{Q\nu R}, \lambda\pi R^2)$ and $\overline{n} = \lambda\pi R^2 - \overline{m}$, with $\nu$ given by $\frac{2rv^2}{(\pi R^2)}$.*

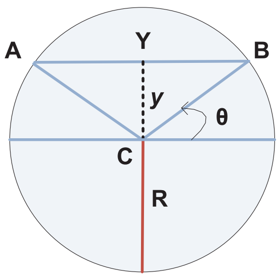


**Figure 7: Chord length in anchor zone**

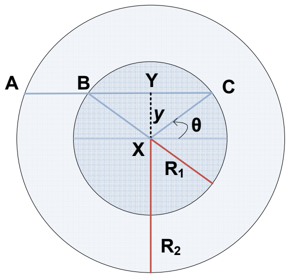


**Figure 8: Anchor zone and ROI.**

PROOF. We denote with $n(t)$ and $m(t)$ the number of nodes with and without content $\mathcal{I}$, respectively, at a given time $t$, and define $N(t) = n(t) + m(t)$. All these quantities vary over time, as nodes move in and out of the anchor zone, and as content is exchanged. We now build a set of differential equation which describe how these quantities vary over time. Consider the time interval $[t, t+\Delta t]$, and let $\delta n(t) = n(t+\Delta t) - n(t)$, and similarly for $\delta m(t)$. $n(t)$ varies over time due to nodes with content $\mathcal{I}$ exiting the AZ, and to nodes without content in the AZ getting the content.

For the first contribution, we assume that users with content are uniformly distributed within the AZ. Then, the average number of users with content in an area $A$ within the AZ is given by $A\frac{n(t)}{\pi R^2}$. Consider now the time interval $\Delta t$: the users with the content that will go out of the AZ are present in the ring of depth $v\Delta t$ around the border of the AZ, whose area can be approximated as $2\pi Rv\Delta t$. On average, half of them have a component of their speed in a direction opposite to the center of the AZ (they are moving out), and half in a direction toward the center (they are moving in). We introduce the approximation that all those nodes in the ring who have a component opposite to the center will leave the AZ in the time interval $(t,\ t+\Delta t)$. Putting all together, $\frac{n(t)v\Delta t}{R}$ is the average number of users with content which go out of the AZ by time $t + \Delta t$.

For the second contribution, the frequency $\nu$ at which two nodes come within the transmission range $r$ of each other inside an area $A$, for the Random Direction mobility model, is given by $\frac{2rv^2}{A}$ [13]. The probability that a node picked at random within the AZ has the content $\mathcal{I}$ at time $t$ is given by $p(t) = \frac{n(t)}{N(t)}$, and the probability that content $\mathcal{I}$ is transferred during an event of two nodes coming into contact within the AZ is $2p(t)(1-p(t))Q$. With $N(t)$ nodes in the AZ, there are $\frac{N(t)(N(t)-1)}{2}$ pairs of nodes that could come into contact. Thus, the average amount of nodes that receive the content $\mathcal{I}$ in the AZ in the considered time interval is given by $\nu\frac{(N(t)-1)(N(t)-n(t))n(t)}{N(t)}Q\Delta t \cong \nu n(t)(N(t)-n(t))Q\Delta t$. Finally, dividing by $\Delta t$ and letting this time interval go to zero, we get

$$\frac{dn(t)}{dt} = \nu n(t)(N(t)-n(t))Q - \frac{n(t)v}{R} \tag{13}$$

The number of nodes inside the AZ without content $\mathcal{I}$ varies in time due to:

- nodes without content $\mathcal{I}$ exiting the AZ,
- nodes entering the AZ (we assume all entering nodes do not have the content, because all nodes delete their copy of $\mathcal{I}$ as soon as they exit the AZ),
- nodes without content $\mathcal{I}$ in the AZ getting the content, by meeting other node(s) with content $\mathcal{I}$.

The first and second contributions can be derived with a similar procedure as used in (13) for computing $n(t)$. Therefore, the average number of nodes without content $\mathcal{I}$ exiting the AZ in the considered time interval is given by $\frac{m(t)v\Delta t}{R}$, and the new nodes entering the AZ in the same time interval is given by $\lambda v\pi R\Delta t$. The third term is the same as in the previous equation, with a change in sign. Putting all together, we get the following differential equation for $m(t)$:

$$\frac{dm(t)}{dt} = \lambda v\pi R - \nu n(t)(N(t)-n(t))Q - \frac{m(t)v}{R} \tag{14}$$

Since we assume that the system is in equilibrium, the time averages of $m(t)$ and $n(t)$, indicated respectively as $\overline{n}$ and $\overline{m}$, remain constant over time. We can write for $\overline{n}$ and $\overline{m}$ differential equations very similar to those derived above, and we can set both $\frac{d\overline{m}}{dt}$ and $\frac{d\overline{n}}{dt}$ equal to zero. Solving for $\overline{n}$ and $\overline{m}$ we get the expressions in the lemma. □

LEMMA 4. *With the same assumptions as before, the probability that a node gets content $\mathcal{I}$ given that it meets $k$ nodes is given by*

$$1 - \left(1 - \frac{Q\overline{n}}{(\overline{m}+\overline{n})}\right)^k \tag{15}$$

PROOF. The probability for a node to successfully get the content upon meeting another node, can be computed by the product of the probability that the encountered node has the content, (equal to the average fraction of nodes having content in the AZ, and given by $\frac{\overline{n}}{\overline{n}+\overline{m}}$), and the probablity of successful information transfer $Q$. The probability in the lemma is then derived as the probability that at least one out of $k$ encounters with other nodes results into a successful transfer of content. $\square$

PROOF. (Result 1) Consider a node traversing the AZ. This node gets content $\mathcal{I}$ if, during its traversal:

- at least one of the encountered nodes has the content, and
- the information is transferred successfully between the two nodes. This implies that the two nodes have been in range for a sufficiently long time to allow the message to be transferred, possibly in presence of transmission errors (collisions and other impairments).

In steady state, the success probability $P_s$ can be written as

$$P_s = \sum_{k=1}^{\infty} P(meet\ k\ nodes) P(get\ \mathcal{I}|\ meet\ k\ nodes) \tag{16}$$

The probability of meeting $k$ nodes is given by Lemma 2, while the probability to successfully get the content upon meeting $k$ nodes is given by Lemma 4. Substituting, we get (1). $\square$

**Proof of Theorem 1:**

PROOF. Let us consider two nodes $A$ and $B$ with transmission range $r$, which come in range of each other. The amount of time these nodes will stay in contact is given by $L_r/v'$, where $L_r$ and $\vec{v'}$ are respectively the length of the chord that one node travels within the transmission range of the other, and their relative speed, with $0 \leq v' \leq 2v$. Using a similar approach as for the derivation of (11), the probability density function of $L_r$ is given by $f_{L_r}(\ell) = \frac{\ell^2}{r^2\pi\sqrt{4r^2-\ell^2}}$. In the RD mobility model, the probability density function of $v'$ is given by $f_{v'}(\omega) = \frac{2}{\pi\sqrt{4v^2-\omega^2}}$. Using (10) and the formula for the pdf of the ratio of two random variables [14], the pdf of the contact duration can be written as (3). If $X'$ is the amount of time required for each transfer attempt, then a necessary condition for the information transfer to be successful is to have a contact duration greater than or equal to $X'$. Including in $S$ all factors relative to communication problems like contention or collisions, we get Eq.2 when nodes continually retry upon transmit failures. $\square$

**Proof of Result 2:** In order to prove Result 2, we first need to introduce the following lemma.

LEMMA 5. *In steady state, for a node traversing an AZ and a ROI, under the Random Direction mobility model, the probability of meeting k nodes before leaving the ROI can be approximated as*

$$P_{BL}(met\ k) = \int_{\sqrt{R_2^2-R_1^2}}^{R_1+R_2} \left( f_{L_{AC}}(\ell) \frac{(2r\ell\lambda)^k e^{-2r\ell\lambda}}{k!} \right) d\ell \tag{17}$$

PROOF. First of all, we compute the average length of the path AC that a node travels inside an AZ and a ROI, before leaving the ROI (as shown in Fig. 9). Let A be the entry point into the AZ, and C be the point where the node exits the ROI, respectively, then, since nodes travel along a straight line, the length of the chord AC is given by

$$L = |AC| = g_2(y) = \sqrt{R_1^2 - y^2} + \sqrt{R_2^2 - y^2} \tag{18}$$

where $y$ is the distance |XY|. Assuming that $y$ is uniformly distributed between 0 and $R_1$, the pdf for the chord length AC can be computed as

$$f_{L_{AC}}(\ell) = \frac{f_L(g_2^{-1}(\ell))}{|g_2'(g_2^{-1}(\ell))|} \tag{19}$$

with $\ell \in [\sqrt{R_2^2 - R_1^2},\ R_1 + R_2]$.

Using a similar approach as in (8), the probability of meeting $k$ nodes is given by (17).

$\square$

PROOF. (Result 2) To compute the success probability before leaving the ROI we can plug (17) and (15) in (16), finally obtaining (4). $\square$

**Proof of Result 3:** In order to prove Result 3, we first need to introduce the following lemma.

LEMMA 6. *In equilibrium state, for a node traversing an AZ and a ROI, under the Random Direction mobility model, the probability of meeting k nodes before entering the ROI can be approximated as*

$$P_{BE}(met\ k) = \int_{R_2-R_1}^{\sqrt{R_2^2-R_1^2}} \left( f_{L_{AB}}(\ell) \frac{(2r\ell\lambda)^k e^{-2r\ell\lambda}}{k!} \right) d\ell \tag{20}$$

PROOF. First of all, we compute the average length of the path that a node travels when it traverses the length $AB$ (as shown in Fig. 9). Since nodes travel in a straight line, the length $L_{AB}$ of the chord AB is given by

$$L = |AB| = g_3(y) = \sqrt{R_2^2 - y^2} - \sqrt{R_1^2 - y^2} \tag{21}$$

The pdf for chord length AB is computed as with $\ell \in [R_2 - R_1,\ \sqrt{R_2^2 - R_1^2}]$.

If the trajectory of a node within the anchor zone is of length $\ell$, the area swept is $2r\ell$, where $r$ is the transmission range of each node, therefore using a similar approach as in (8), the probability of meeting $k$ nodes along the chord AB is given by (20).

□

PROOF. (Result 3) Using equations (20) and (15) in (16) we finally get (6). □