On maximizing the lifetime of distributed information in ad hoc networks with individual constraints

C. Westphal

***Abstract*— Ad-Hoc networks and in particular sensor networks are networks of nodes with limited battery and limited processing capacity. As such, a single node carries an incentive to limit the amount of data it contains. This leads to the expiration of the data carried by the node after a period of time, due for instance, to a re-boot after an off period in the duty cycle, or to older information being "pushed" out by new data received by the mobile node. On the other hand, some data is critical to the functioning of the whole network. For instance, the existence and position of a gateway towards the infrastructure network should be kept *somewhere* in the network, so that nodes are able to recover this information when needed. In this extended abstract, we study the trade-off between the finite lifetime of a piece of information at each node, and the survival of this information indefinitely within the network. We consider a simple dissemination process for the information akin to an AODV-based information request/reply mechanism. We show that the maximum number of hops in a request is a critical parameter to ensure the survivability indefinitely of any information within the network. We identify the parameter which minimizes the load on the network, for a network graph satisfying the Poisson boolean model. We also show how to minimize the cost of the dissemination on the network, so as to keep this cost decreasing asymptotically to 0.**

***Index Terms*— Ad-hoc networks, sensor networks, contact process, peer-to-peer systems.**

I. Introduction

AD-HOC networks are collection of nodes organized without the assistance of an infrastructure. The nodes join and leave the network, or move about, making it a permanent challenge to identify which node is able to provide which information to the other nodes in the network.

Further, resource constraints on the node limit the amount of data carried by each node. Due to these resource constraints, the data carried by the node disappears, or expires after some time. For instance, in a sensor network, a node might alternate between on and off states to economize battery power. At every wake-up time in the duty cycle, the states that were maintained in the previous cycle might have been flushed out. The node then starts the new cycle with a blank slate. The lifetime of the data which composed these states is thus that of the new *on* cycle.

In another example, the constrained resource could be the memory storage itself: a node in an ad hoc network might store data that keeps streaming in. For instance, a set of nodes sharing mp3 or mpeg files in an ad-hoc peer-to-peer network: the capacity of the typical portable multimedia player is much less than the size of a typical music library, let alone a movie library. After some initial period filling in the memory, some choice has to be made as to what to keep and what to discard. If all the data has the same priority, then a FIFO policy could be used to discard the oldest data and replace it with the newest. Other policies would yield different length of stay for the data on the device, but in most case, a finite lifetime is assigned to the data by the user's policies and resource constraints.

In an ad-hoc networks, without even considering that nodes could move relative to each other, locating data objects which has a finite lifetime might prove difficult: the data object that existed at time t might have expired at time $t + \tau$. Since each single piece of data will eventually disappear, the data consisting of a given object must be replicated throughout the network. For instance, a peer-to-peer network will contain multiple copies of the same mp3 file. Multiple sensor nodes in a wireless sensor network will contain a network configuration file pointing to the physical location of the information sinks in the sensor network, or some routing information, or some other data object pertaining to the function of the network.

While each node tends to discard data objects to satisfy its resource constraints, the network as a whole has an incentive to ensures that copies of each object are alive and distributed throughout the network. The object might be necessary for the network to work properly, as in the configuration file example in a sensor network. Or the wide availability of even the rarest objects might be an incentive for nodes to join the network, as in the peer-to-peer example. The network incentive is thus to preserve the existence of data objects while each single node has the resource constraint to remove such objects.

A. *Dissemination Process*

The issue we are interested in becomes dual: first, is it possible to ensure that *live* replicas of a given object exist within the network at all times? And second, when a node requires the use of a data object, is it possible to locate such object efficiently?

Before stating the main results, we consider the problem of locating objects distributed throughout the network. A node needs to access a given object. In an ad hoc network, this first node is unable to refer to a centralized location registry, as there is no such infrastructure. The simplest way is for this node to broadcast an object request. The request is sent to its neighbors, who relay it to their neighbors, etc., until it finds a second node that has a live copy of the object. This second node can then reply to the first one with a copy of the object. This is the principle that AODV [14] uses to locate the correspondent of a node for routing purpose. It has been suggested [11] to extend this protocol to locate a more general class of objects than routing data. We consider a similar request/reply mechanism for the problem of locating objects in the ad hoc network.

The request/reply mechanism is attractive as it both locates and replicates the data objects. Each positive reply creates a fresh copy of the object at the requestor, which then uses the object for its intended purpose and stores it for other for a period of time. We call this mechanism the *dissemination process*.

More sophisticated solutions using distributed hash tables (see for instance [15], [16] and numerous others) exist, but we do not consider them at this point: since the lifetime of the object is not known *a priori* but is rather a parameter of the node holding the object, it is difficult for a DHT to always point to a live copy. If the DHT does not point to a live copy, then the broadcast mechanisms becomes the default mode anyway.

Of course, flooding the network will find any copy of the requested object. However, we would like to minimize the impact of the broadcasting phase, mostly by limiting the broadcast radius. As the number of nodes grows, the amount of broadcast traffic grows as well. Mechanisms exist to reduce the amount of data (minimal spanning tree broadcasting mechanism, such as [1] for instance) but they add on processing complexity.

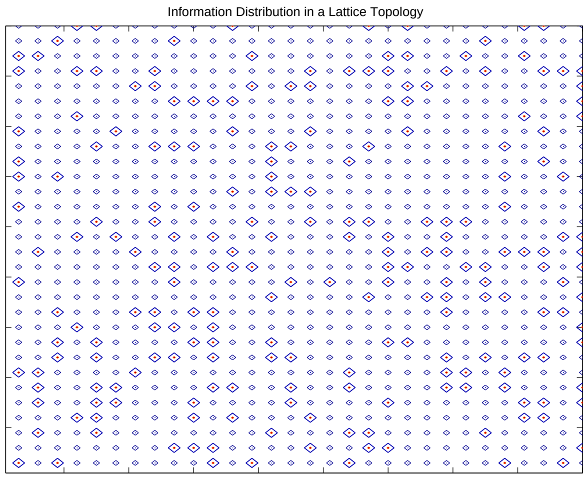

Fig. 1. Information distribution over a lattice

Figure 1 depicts a lattice network which was uniformly seeded with replicas of one object. It shows the distribution of the objects after running the reply/request dissemination process: nodes with a thicker diamond mark are the nodes which store the valid object, nodes with a light mark do not maintain a state with the object. However, the object is distributed somewhat uniformly in this scenario, and it is available to most nodes within 3 hops.

A contribution of the paper is that it is the first study of an ad hoc network that stores object with a finite lifetime. The typical assumption in distributed systems is that the information is not perishable, something that is not true due to the individual nodes constraints in an ad-hoc or a sensor network. We also believe that it is one of the first attempts at generalizing the contact process to the field of ad hoc networks, where we believe it would find a wide applicability. It has been used in a communication context, but in one dimension by [17]. This paper extends the results of [18]. [18] initiates the analysis presented here, but focuses on two topologies for ad hoc networks: the tree topology and the lattice topology. These are reasonable topologies to study: ad hoc routing in some network is done according to a minimal spanning tree, and the tree structure is thus widely used. Lattice topologies are used to efficiently cover areas. This work extends the results in [18] to random topologies.

We now detail the analytical model before stating the results.

II. Network Model

We consider the network graph as a Poisson boolean model $\mathcal{B}(\zeta, R)$, where the nodes are distributed according to a Poisson point process in $\mathbf{R}^2$ with intensity ζ and two nodes are connected if they are within distance R of each other. We set R to be supercritical, as no dissemination can occur otherwise.

III. The Threshold Contact Process

The threshold contact process was studied by [9]. It is a variant of the contact process first described by Harris [10]. For an in depth description of the contact process, we refer the reader to [12], [13]. The threshold contact process is a Markov process on the lattice.

In 2 dimension, the process $\{X_t(v)\}$ takes value 1 if the node v with coordinates $(x, y) \in \mathbf{Z}^2$ possesses some given object, or 0 otherwise. We will write that a node v is infected if $X(v) = 1$, or that the site v is occupied.

In the threshold contact process, an occupied node becomes vacant with rate 1, while a vacant node becomes occupied with rate λ if at least θ nodes $w \in v + \mathcal{N}$ are occupied. $\mathcal{N}$ is a convex finite subset of the lattice centered at the origin, which is invariant by the reflections about and the permutation of the coordinate axes. In our case, we consider a particular case of the threshold contact process, with $\theta = 1$ and $\mathcal{N}$ the set of nodes at distance D of the origin, where the distance of (x, y) to the origin is $|x| + |y|$.

The initial condition of the process is the finite set S of nodes such that, at time 0, $X_0(s) = 1$ if $s \in S$ and 0 otherwise.

One quantity of interest is the extinction time for the contact process when the initial condition at time 0 is S as it depends on λ, namely $\tau_S(\lambda)$:

$$\tau_S(\lambda) = \inf\{t : \forall x, y \in \mathbf{Z}^2, X_t(x, y) = 0\} \qquad (1)$$

One canonical result is that there exists a value of $0 < \lambda_c < \infty$ which defines a critical threshold: if λ is less than λ_c, then λ is described as subcritical, and:

$$P[\tau_S(\lambda) < \infty] = 1 \qquad (2)$$

Otherwise, if $\lambda > \lambda_c$, that is, in the supercritical case:

$$\Theta(\lambda) \triangleq P[\tau_S(\lambda) = \infty] > 0 \qquad (3)$$

In the first case, the process dies out in a finite time almost surely. In the second case, there is a positive probability that the process never becomes extinct. While the second case does not guarantee the survival, the first case guarantees the extinction: this should be enough motivation to identifying the value which separates both cases.

Θ can be made close to 1 by choosing the right λ. Whenever we talk about an object staying alive indefinitely, it is to be understood: staying alive with some chosen probability Θ.

IV. Criticality of the Boolean Model

The first contribution of this paper is to generalize the threshold contact process to the Poisson boolean model, and to identify the existence of a critical radius D. For a fixed λ, there exists a radius D_c such that, for $D > D_c$, $P(\tau = \infty) > 0$.

We now state the results formally.

Lemma IV.1: For a fixed λ, there exists a critical radius $D_c^{\mathbf{Z}^2}$ for the threshold contact process on the lattice $\mathbf{Z}^2$

For $D < D_c$, the process is subcritical, and, if $\tau(D)$ is defined as, for some finite initial condition set S:

$$\tau_S(D) = \inf\{t : \forall x, y \in \mathbf{Z}^2, X_t(x, y) = 0\} \qquad (4)$$

then:

$$P[\tau_S(D) < \infty] = 1 \qquad (5)$$

Otherwise, if $D > D_c$, that is, in the supercritical case:

$$\Theta(D) \triangleq P[\tau_S(D) = \infty] > 0 \qquad (6)$$

Proof: We restrict ourselves here to the processes on the lattice.

Increasing D increases the number of nodes that can request the object from a given node, and thus the likelihood that a node will disseminate the object. The system with λ fixed and a variable D can be mapped to system with D fixed and a variable λ. We need to show that the critical behavior of λ translates into a critical behavior for D.

We define by $S_{k,l}$ the square composed of the nodes of coordinates $kD + i, lD + j$ where $i, j \in \{0, \ldots, D-1\}$. We define the process $Y_t(k,l)$ to be equal to 1 if $X_t(v) = 1$ for at least one $v \in S_{k,l}$ and 0 otherwise. The process Y is an aggregation of X over the squares $S_{k,l}$.

We start the process Y at time 0 and define a process Y' that is equivalent to Y with the following difference: if $Y_t(k,l)$ turns from 0 to 1 by receiving the object reply from a node in either $S_{k+1,l+1}, S_{k-1,l+1}, S_{k+1,l-1}, S_{k-1,l-1}$, then Y' stays equal to 0. This means that $Y'_t(k,l)$ can be infected only by its single hop neighbors on the (k,l) lattice, not by the nodes that are diagonally connected to it. $Y' \leq Y$ by construction, for all t and all (k,l).

Now, we also assume that Y' goes from 1 to 0 according to the process of one of the infected node in $S_{k,l}$. Which one does not matter as they are all distributed independently with a Poisson process with the same rate. Again, Y' goes to 0 faster than Y, since Y will turn to 0 only when all the nodes in the square have a value 0 for X (including the one we chose for Y').

The rate of requests in the process Y' from its single hop neighbor is $D^2\lambda$ (we consider only the request that infect the neighbor). Thus Y' is a threshold contact process with radius 1, extinction rate 1 and rate $D^2\lambda$. We know that $\infty > \lambda_c^1 > 0$ exists for such process, and thus, taking D big enough such that $D^2\lambda > \lambda_c^1$, the process Y' is supercritical.

This in turn implies that Y is supercritical, and X must be as well. ∎

Theorem IV.1: For a fixed λ, there exists a critical radius $D_c^{\mathcal{B}}$ for the threshold contact process on $\mathcal{B}(\zeta, R)$.

Proof: To construct the lattice, pick a length l, and define the square $S(i,j)$ to be the set of points (x,y) such that $il \leq x < (i+1)l$ and $jl \leq y < (j+1)l$. Since the Poisson boolean model is supercritical by assumption, we can pick l big enough such that, with probability p there is a path within $S(i,j)$ connecting the left to the right to the top to the bottom sides, a path within $S(i, j+1)$ connecting the left to the right to the top to the bottom, and both paths are such that they connect at the boundary between $S(i,j)$ and $S(i, j+1)$. Symmetry ensures that a similar path exists, rotated 90 degrees, between $S(i,j)$ and $S(i+1, j)$ with the same probability p.

We now consider the bond percolation model on the lattice and say that the edge between (i,j) and $(i, j+1)$ (and similarly $(i-1, j)$, $(i, j+1)$ or $(i, j-1)$) is open if there exists the ”Lorraine cross” shaped path described above in the Poisson boolean model joining $S(i,j)$ to $S(i, j+1)$. This defines a bond percolation model with edge probability p. For a similar construction, see [2].

Pick some $L \in \mathbf{Z}$. Denote by $\sum(v,w)$ the square in the lattice containing the points (i,j) such that $vL \leq i < (v+1)L$ and $wL \leq j < (w+1)L$. Define by A_L the event that there is a ”cross of Lorraine” shaped path joining two neighboring $\sum$ squares in the bond percolation model.

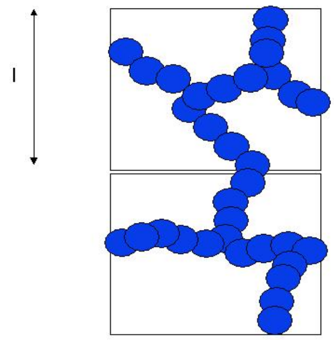

Fig. 2. Cross of Lorraine path joining to neighboring squares

By a well known result (see for instance [8] Lemma 11.22 extended to our case), we have for some $\alpha, \gamma > 0$:

$$P_p(A_L) \geq 1 - \alpha e^{-\gamma L} \tag{7}$$

Thus, we probability $p_L = P_p(A_L)$, a request issued by a node in the connected cluster in one square will be forwarded to each of its neighboring squares. Denote by a hop the transmission of a request from inside one square $\sum$ to a neighboring square.

A request set to travel Δ hops in the $\sum$ lattice will travel on average at least $p_L^{\Delta}\Delta$ hops. By a sphere packing argument, each hop is composed of at most $O(L^2)$ hops within each $\sum$ square times at most l^2/R^2 hops within each S square. If we set l for a given p, then the number of relays for a request is of the order of $O(L^2)$.

Thus we can write the relation of p_L^Δ as a function of L:

$$p_L^\Delta \sim (1 - \alpha e^{-\gamma L})^{L^2} \rightarrow_{L\rightarrow\infty} 1 \quad (8)$$

Thus we can pick L large enough and Δ large enough so that $p_L^\Delta \Delta > D_c$ as per Lemma IV.1. These values ensures the survivability of the threshold contact process in the Poisson boolean model.

Putting everything back together, we see that setting the number of hops in the Poisson boolean model to be more than:

- l^2/R^2 hops in the square S times
- L^2 hops in the square $\sum$ times
- Δ hops in the larger lattice

Taking $D > \Delta L^2 l^2/R^2$ yields survivability of the process. We have proved that there is a value for which the process survives indefinitely with some strictly positive probability, thus we showed the existence of a critical radius for the threshold contact process in the Poisson boolean model ∎

V. Cost of the dissemination

The previous section dealt with disseminating information in the network so as to ensure it survives. We have seen that if each nodes requests the information from its neighbors D hops away, then the information is preserved in the network, with some probability Θ.

The cost in number of messages per node per unit of time can be assessed. It depends on the number of neighbors of each node, which depends on the intensity ζ of the Poisson process in the Poisson boolean model, and on D. It is of the order of $\lambda(\zeta\pi R^2)^D$.

The next theorem shows how to minimize the cost by dividing the nodes in three categories: nodes which request and store the information, nodes which only forward the information, and nodes which stay idle.

Theorem V.1: Define the cost c of maintaining the information alive to be the number of messages transmitted per node per unit of time. For each $\epsilon > 0$, we can construct a dissemination mechanism such that $c < \epsilon$.

Proof: We consider again the squares $S(i,j)$ as per the previous section. We choose one node on the connected cross of Lorraine to be a request-and-store node in each of the square. We say these nodes are of type RS. Between two such nodes in neighboring squares, by our construction, there is a path joining these nodes with probability p. The nodes on this path are forwarding nodes, or type F: these nodes only carry the requests and reply between the nodes of the type RS. All the other nodes are idle nodes, and do not participate in the dissemination.

This defines a lattice overlay network over the original network. We need to assess the number of hops K between two nodes of type RS, or some upper bound thereof. Each request-and-store nodes are at K hops away, so a request sent over KD hops will visit D request-and-store nodes. The threshold contact process can be kept alive indefinitely with probability Θ, for $D > D_c^{\mathbf{Z}^2}$.

To assess the cost, see that there are about ζl^2 nodes in the square, but transmission involves at most $K4.3^(D-1)$ transmission. This upper bound comes from the fact that each square other than the origin can forward the request to at most 3 neighbors. Using lemma V.1, which we will detail below, and taking l large enough ensures that $c < \epsilon$. ∎

Lemma V.1: Define by K the length of the path, measured in hop count, from one request-and-store node to the next. $K = o(l^2)$

Proof: The intuition behind this result is that one can narrow a path into a smaller and smaller area of the square as the size of the square grows larger.

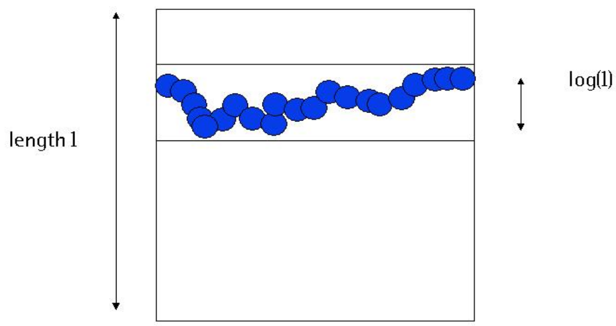

Fig. 3. Logarithmic bounds on the crossing paths

A consequence of Theorem 3 in [6] is that each of the vertical and horizontal crossing which connect the nodes of type RS can be found with high probability into rectangle of size $l \times \log l$. The number of relay in between two RS nodes is thus bounded by the area of such narrow rectangles over the size of the square.

By taking l large enough, we ensure that the probability of finding path within the $l \times \log(l)$ rectangle with high enough probability $1 - \epsilon/2$. The cost of disseminating the information becomes bounded by a quantity of the order $\log(l)/l$. Taking l such that $\log(l)/l < \epsilon/2$ gives a total forwarding cost bounded above by ϵ. ∎

VI. Conclusion

We have shown that the threshold contact process on the Poisson boolean model exhibit a phase transition for the broadcast radius D, and that a reply-request can be used to keep the information objects alive within the network at an arbitrarily low cost. Further issues to consider involve the design of the protocol to ensure that all the nodes that are part of the infinite cluster in the Poisson boolean model take turn being a request-and-store node, a forwarding node or an idle node.

Another issue of interest is how to reduce the cost of disseminating the information while keeping it available to all the nodes with some probability, and not only to an overlay subset. Hierarchical schemes spring to mind, and this is the focus of future research.

The infinite size of the network runs against practical situations. We would like to study the asymptotic behavior of our dissemination process as the size of the network grows. The contact process shows different asymptotics above and below a critical threshold [3], [4], [5], and it would be interesting to study the behavior of the dissemination process in this context.

Acknowledgments

The author would like to thank Aaron Wagner and Venkat Anantharam for their invaluable and helpful discussions.

References

[1] Chen B., Jamieson K., Balakrishnan H., Morris R. *Span: An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks*, ACM Wireless Networks Journal, Volume 8, Number 5, September, 2002. Pages 481 - 494.

[2] Dousse O., Mannersalo P., Thiran P. *Latency of wireless sensor networks with uncoordinated power saving mechanisms* in Proc. of MobiHoc 2004, Tokyo, Japan.

[3] Durrett R., Liu X., The contact process on a finite set Ann. Probab. 16 (1988), 1158-73

[4] Durrett R., Schonman R., The contact process on a finite set II Ann. Probab. 16 (1988), 1570-83

[5] Durrett R., Schonman R., Tanaka N., The contact process on a finite set III. The critical case Ann. Probab. 17 (1989), 1303-21

[6] Franceschetti M., Dousse O., Tse D., Thiran P. *On the throughput capacity of random wireless networks* Preprint.

[7] Gilbert E. *Random Plane Networks* J. Soc. Industrial Applied Mathematics, Vol. 9, No. 4, 1961, pp.533-43.

[8] Grimmet G. *Percolation* Springer-Verlag, New York, 1991.

[9] Handjani S. *Survival of Threshold Contact Processes* Journal of Theoretical Probability 10 (3): 737-746, July 1997.

[10] Harris T.E. *Contact Interactions on a Lattice* in Annals of Probability, 1974, Vol.2, No. 6, 969-988.

[11] Koodli R., Perkins C. *Service Discovery in On-Demand Ad Hoc Networks* IETF Manet Working Group, Internet Draft draft-koodli-manet-servicediscovery-00.txt, 9/2002.

[12] Liggett T. *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes* Springer-Verlag, New-York, 1999.

[13] Liggett T. *Interacting Particle Systems*, Springer-Verlag, New York, 1985.

[14] Perkins C., Royer E. *Ad hoc On-Demand Distance Vector Routing.* Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, February 1999, pp. 90-100.

[15] Plaxton C., Rajaraman R., Richa A. *Accessing Nearby Copies of Replicated Objects in a Distributed Environment* ACM Symposium on Parallel Algorithms and Architectures, 1997.

[16] Stoica I., Morris R., Karger D., Kaashoek F., Balakrishnan H. *Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications* in Proceedings of ACM SIGCOMM, 2001.

[17] Wagner A., Anantharam V. *Designing a Contact Process : The Piecewise-Homogeneous Process on a Finite Set with Applications.* Stochastic Processes and their Applications (to appear, available from http://www.eecs.berkeley.edu/ ananth/).

[18] Westphal, C. *On maximizing the lifetime of distributed information in ad hoc networks with individual constraints: lattice and tree topologies* to appear in Proc. of Mobihoc 2005, May 2005, Urbana-Champaign, IL.