Study and Design of Algorithms for Information Dissemination in Unstructured Networking Environments

Dimitrios Kogias*

Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
Greece
Email: dimkog@di.uoa.gr

Abstract. The focus on this thesis lies on the study of several information dissemination techniques in modern unstructured networks. Information dissemination has an important academic interest in these environments due to the special characteristics (e.g., decentralization, large-scale and dynamic nature) that they possess.
One of the techniques that is studied is *probabilistic flooding* and analytic (asymptotic) bounds on the value of the forwarding probability p_f, for which a probabilistic flooding network manages to fully cover an underlying connected random graph are presented. The technique of *multiple random walkers* is, also, studied given analytical expressions regarding coverage and termination time for fully and less dense connected topologies. The observation that the network is not efficiently covered at the early stages due to the (potentially) large collection of walkers at the initiator node has led to the introduction of a new information dissemination mechanism that creates walkers (by *replicating* the existed ones) during the random walks and not from the beginning of the process. This replication technique, called *Randomly Replicated Random Walkers (RRRW)*, has been studied in various networking topologies (e.g., random geometric graphs, power- law graphs, clustered graphs) to examine whether it can fill the gap between the performance of two well-known techniques, the *Full Flooding* and *Single Random Walker (SRW)* in *stretching* the advertising information in broader networking areas.

1 Introduction

In modern networking environments, the discovery of a given piece of information plays a key role to its robustness and functionality. In particular, easy and quick access to any information source that is needed is expected, along with the possibility of sharing it with more network nodes located in further network areas.

In this thesis, a study of various information dissemination techniques (existed and newly introduced ones) will take place in unstructured networking

* Dissertation Advisor: Ioannis Stavrakakis, Professor

environments. These environments are mainly distributed topologies, where all the nodes are equal and are characterized by their large size, their scalability properties and their highly dynamic nature. Due to these inherent characteristics, it is not possible for a node to possess information regarding the global network structure in any given time. On the other hand, a node will always know the number and the identity of its one hop neighbours. This lack of knowledge, regarding the structure of the overall topology (even though it exists), by any network node is the reason for calling these networks *unstructured.* P2P and ad-hoc networks possess the described characteristics and will be, therefore, considered and studied throughout this thesis.

Because of the aforementioned characteristics, the process of disseminating information in an unstructured environment becomes very difficult and costly (both in number of messages and in termination time) but, in the same time, it is the reason for its high academic and research interest. In this thesis, the information that is disseminated is related to the knowledge of the location in the network of a node that possess a certain service. Therefore, it is considered as part of a larger process that is called *Service Discovery.*

The process of Service Discovery will be divided in two phases: *Service Advertising* and *Service Searching*, that are sequentially applied. During service advertising, a node advertises (using an information dissemination technique) his location, along with the service he possesses, to part (or even whole) of the network by constructing an *advertising network* (i.e., a network consisted by the nodes and links used for advertising). During searching, the node first checks whether he is part of the advertising network of the desired service and if he is then the searching is successfully completed. If not, then he employs an information dissemination technique to search for either a node that is part of the advertising network or the node that possesses the service itself.

It can be easily verified that this two service discovery phases are complementary in nature and that when the intensity of one is small then the intensity of the other should be large.

1.1 Related Work

Various techniques have been proposed so far for disseminating information in a network. The most popular ones will be presented here.

The simplest technique used for information dissemination (both for advertising and for searching) is *traditional flooding*, [1]. Under traditional flooding, the information messages traverse all network links and, thus, visit all the nodes in the network, producing a large number of messages, especially when the network's size (i.e., N) increases. *Termination time*, on the other hand, is significantly small, upper bounded by the network *diameter*, typically of the order of $log(N)$.

A popular variation is the *controlled flooding* technique, which employs flooding but only for a number of K hops (i.e., K is a Time- to- Live, TTL, value) away from the node that initiates the process (called the *initiator node*). While the value of K remains small, the number of messages produced also remains small and so does the size of the network covered by this technique, reducing the probability for efficient discovery of the desired node or service in a large-scale, modern environment. The controlled flooding (or *K- flooding*) technique is known to be used for searching in the Gnutella P2P system, [2].

On the other hand, techniques like *random walks*, [3], are very different than flooding. These approaches manage to reduce the total number of its messages by sending a limited number (one in the case of the *Single Random Walker* or m in the case of *Multiple Random Walkers*) of entities (as special messages) to cover the network. Each of these entities follows its own path by randomly selecting the next node to visit from one of the 1- hop neighbours of the node that the entity resides in each time slot. The termination of the algorithm takes place after some predefined time (e.g., using TTL expiration), for each entity, or after checking with the initiator node and learning that the desired information has already been discovered by another entity. A combination of the aforementioned termination conditions can also be applied.

Hybrid probabilistic techniques (e.g., local flooding process initiated after a random walk) have also been proposed and analysed, [4], as well as other schemes that adapt the employed TTL values in a probabilistic manner, [5]. Another modification, [6], allows for network nodes to forward messages to their neighbours in a random manner, thus significantly reducing the number of messages in the network. Many other works have been published proposing the selective forwarding of a certain message in the network, e.g., [7, 8].

As it has already been mentioned the main scope of this thesis is the study of several existed and newly introduced information dissemination techniques. Therefore, the rest of this work is organized as follows: in Section 2 the study of the *probabilistic flooding* technique will be presented and an asymptotic analysis following the bounds on the value of the forwarding probability p_f will take place. In Section 3 the technique of *multiple random walkers* is studied giving analytical expressions regarding coverage and termination time for fully and less dense connected topologies. For more efficient coverage of the network during the early stages, a new information dissemination mechanism that creates new entities (by *replicating* the existed ones) is introduced. In Section 4 the technique of *Randomly Replicated Random Walks (RRRW)*, is studied in various networking topologies (e.g., random geometric graphs, power- law graphs, clustered graphs) to examine whether it can fill the gap between the performance of two well-known and used techniques, the *Full Flooding* and *Single Random Walker (SRW)* in stretching the advertising information in wider network areas.

2 Probabilistic Flooding

When *probabilistic flooding* is applied every node, by receiving the message for the first time, forwards it to each of his 1 hop neighbours (apart from the one(s) that forwarded it to him) using a constant forwarding probability $p_f(N)$. By properly parametrized the value of this probability $p_f(N)$ it is possible to cover the underlying graph while producing a smaller number of messages than the traditional flooding approach. The cost to be paid for this reduction is that the coverage is no longer deterministically guaranteed but is rather probabilistically achieved.

Each time probabilistic flooding is applied to a graph, a *probabilistic flooding network* is created. This is a connected network which contains the number of nodes and links over which the disseminated information has been forwarded. The main scope for using probabilistic flooding is the creation of a probabilistic flooding network (i.e., $P(G(N,p),p_f)$) that contains the minimum number of links/ messages, while still achieving the desired network coverage.

In this thesis, the use of probabilistic flooding is studied when the underlying graph is a random graph. In fact, it is noticed that there is a connection between the stages for the creation of a random graph $G(N,p)$ (which follows the binomial model and, thus, has an expected number of links that equals $p(N)\frac{N(N-1)}{2}$, [9]), depending on the value of $p(N)$, and the stages for the creation of the probabilistic flooding network, depending on the value of $p_f(N)$, when probabilistic flooding is applied to a connected random graph. The observation of such a connection allowed for the a further study of probabilistic flooding with the use of elements from graph theory.

2.1 Analytic bounds on p_f for global outreach of $G(N,p)$

One of the problems that has been studied was to find analytic (asymptotic) bounds for the forwarding probability $p_f(N)$, in order to achieve full coverage of an underlying random graph $G(N,p)$, while producing the minimum number of information dissemination messages. In order to achieve this, the use of two random graphs $G(N,p*p_f)$ and $G(N,p*p')$ is proposed. This two graphs have the same value of p with the one that is used to create the connected random graph $G(N,p)$.

Since a link of the underlying connected random graph $G(N,p)$ will be part of the probabilistic flooding network either with probability $p_f(N)$ (when only one of the end nodes of the link receives the message from an other link and takes a forwarding decision) or with probability $p'(N) = 2p_f(N) - p_f^2(N)$ (when both the end nodes of the link receive the message from a different than their common link and, thus, both make an (assumed independent) decision to forward it over the common link), it is expected that $P(G(N,p),p_f)$ contains on average more links than $G(N,p*p_f)$. Consequently, when $G(N,p*p_f)$ is connected with high probability (i.e, *w.h.p* from now on), then $P(G(N,p),p_f)$ is also connected *w.h.p* and, thus includes all network nodes *w.h.p.* Note also that since $p_f(N) \leq p'(N)$ (the equality holds for $p_f(N) = 1$), $G(N,p*p')$ contains (on average) more links than $G(N,p*p_f)$ and when the latter network is connected the former is also connected *w.h.p.*

Based on the two previous observations, and assuming a certain value for $p(N)$, as $p_f(N)$ increases, it is expected that there will be some probability value for $p_f(N)$ for which $G(N,p*p')$ becomes connected *w.h.p.* As $p_f(N)$ increases further, $P(G(N,p),p_f)$ becomes connected (equivalently, $C(0) = 1$) *w.h.p.* For further increment, $G(N, p * p_f)$ also becomes connected. Consequently, the particular value of $p_f(N)$ for which probabilistic flooding disseminates information to all network nodes (thus achieve global network outreach) is "between" the values of $p_f(N)$ for which $G(N, p * p')$ and $G(N, p * p_f)$ become connected. This analysis has been also verified by simulation results in a $G(10000, 0.0008)$ random graph. The simulation results have also shown that this behaviour is also present when smaller network coverage cases are studied, something that has not been covered by the analysis.

An interesting result is that even though $G(N, p * p')$ becomes connected for smaller values of $p_f(N)$ when compared to $G(N, p * p_f)$ (as already mentioned $p' > p_f$ for $0 < p_f < 1$), these values have the same asymptotical order.

2.2 Probabilistic Versus Full Flooding

To be able to study cases of smaller coverage (than the global outreach of the underlying connected graph), a new metric was introduced.This metric is the $L - coverage$: it includes the number of nodes that have been informed along with those nodes that are at most L hops from (at least) one of them. The cases of $L = 0$ (i.e., global network outreach) and $L = 1, 2$ are covered here. Larger values for L were not studied due to the *small world phenomenon*.

The next problem that was studied was a comparison of the probabilistic flooding approach with the traditional flooding. A reduction on the number of messages to be achieved under probabilistic flooding at a cost of an increase in the termination time until global outreach of the underlying random graph was expected and verified both by analytic and by simulation results.

Let $R_{M,L}(N)$ denote the (asymptotic) fraction of messages under probabilistic flooding over those under traditional flooding for some L, or, $R_{M,L}(N) = \frac{\mathcal{E}(\mathbb{P}(\mathbb{G}(N,p),p_f))}{\mathcal{E}(\mathbb{G}(N,p))}$. For the case of $L = 0$,

$$R_{M,0}(N) = \frac{\ln(N)}{p(N)N}. \tag{1}$$

Obviously, $R_{M,0}(N) \to 0$, when $N \to +\infty$ ($O(p(N)) > O(p_{Q_0}(N)) = O(\frac{\ln(N)}{N})$). Note that in strict terms, $R_{M,0}(N) = \Theta\left(\frac{\ln(N)}{p(N)N}\right)$.

As already mentioned, since $\mathbb{G}(N, p)$ is a connected network w.h.p., then $O(p(N)) > O(p_{Q_0}(N)) = O(\frac{\ln(N)}{N})$. For $p(N) = \Theta\left(\frac{\ln(N)}{N}\right)$ (which means that $\mathbb{G}(N, p)$ has just become connected), it is interesting to see that $R_{M,0}(N) = \Theta(1)$, which apparently demonstrates the fact that there is no advantage under probabilistic flooding when compared to traditional flooding for this case (the number of messages is the same under both probabilistic flooding and traditional flooding). Actually, this particular case is -asymptotically- equivalent to $p_f(N) = 1$, for which probabilistic flooding reduces to traditional flooding. In order to explain further this observation, note that for the case of $p(N) = \Theta\left(\frac{\ln(N)}{N}\right)$, $\mathbb{G}(N, p)$ has just become connected w.h.p. which apparently means that the number or “redundant” links (links over which traditional flooding forwards messages and probabilistic flooding “saves” by probabilistically “avoiding” to do so) is significantly reduced. The shape of the network – even though it contains cycles – looks mostly like a tree, and therefore, the ability of probabilistic flooding to “avoid” forwarding messages over “redundant” links is reduced.

The reduced number of messages under probabilistic flooding is achieved at the expense of larger termination delays. This is shown by comparing the network diameter of $\mathbb{G}(N,p)$ and $\mathbb{P}(\mathbb{G}(N,p),p_f)$, for $p_f(N) = \Theta\left(\frac{\ln(N)}{N}\right)$ (the upper bound of termination time corresponds to the network diameter). Let $R_{T,L}(N)$ denote the (asymptotic) fraction of the network diameter of the probabilistic flooding network over the network diameter minus L (for fairness issues), for some $L = 0, 1, 2$, or $R_{T,L}(N) = \frac{\mathcal{D}(\mathbb{P}(\mathbb{G}(N,p),p_f))}{\mathcal{D}(\mathbb{G}(N,p))-L}$. Given that for $L = 0$, $\mathcal{D}(\mathbb{G}(N,p)) = \Theta\left(\frac{\ln(N)}{\ln(p(N)N)}\right)$, $\mathcal{D}(\mathbb{P}(\mathbb{G}(N,p),p_f)) = \Theta\left(\mathcal{D}(\mathbb{G}(N,p*p_f))\right) = \Theta\left(\frac{\ln(N)}{\ln(p(N)p_f(N)N)}\right)$, and $p(N)p_f(N) = \Theta\left(\frac{\ln(N)}{N}\right)$, it follows that,

$$R_{T,0}(N) = \frac{\ln(p(N)N)}{\ln\left(\ln(N)\right)}. \tag{2}$$

So far, the global network outreach case (i.e., $L = 0$ or $C(0) = 1$) has been studied. The cases corresponding to $L = 1$ and $L = 2$ are naturally expected to yield smaller number of messages under probabilistic flooding – compared to traditional flooding – since the particular values of $p_f(N)$ are expected to be (on average) smaller than those ensuring global network outreach (i.e., $C(0) = 1$) w.h.p. The asymptotic analysis that has been previously followed for $L = 0$ applies to these particular cases as well. Note, that the resemblance is only asymptotic and savings with respect to the number of messages are greater for the case of $L = 1$ and $L = 2$ than for the case of $L = 0$ under probabilistic flooding.

Simulation results regarding the number of messages and termination time for $L = 0, 1$ and 2 have verified the aforementioned analysis.

3 Multiple Random Walker

In [10] it is shown that *multiple random walkers*, starting from the same network node, are capable of accelerating the information dissemination process and reduce termination time by a factor equal to the number of random walkers, for a wide range of topologies. On the other hand, as the number of random walkers increases, the number of messages sent increases proportionally to the number of random walkers. Moreover, since random walkers start from the same network node, it is expected for some initial movements to partially overlap (thus, not improving coverage) due to visits to already visited network nodes (i.e., *revisits*). This motivates the adoption of a *replication* approach -under which replicas of random walkers are probabilistically created after each movement- so as to avoid initiating all of them at the same time as under the multiple case. A simple replication mechanism is proposed here capable of covering larger network areas than multiple random walkers for the same number of random walkers and allowed number of messages.

The contribution of this thesis is the study of multiple random walkers from a different perspective than the one presented in [10]. The analytical part of the work initially assumes a fully connected network topology which allows for the derivation of an analytical expression that confirms the results presented in [10], allowing also for further understanding of various aspects of information dissemination under multiple random walkers. The analysis continues capturing coverage in less dense topologies and an analytical expression is derived showing how coverage is affected by frequent random walk revisits.

3.1 Multiple Random Walkers

Having started with m random walkers from the same initiator node, each random walker moves to one of its neighbour nodes being selected randomly and independently among the set of neighbour nodes, provided that the chosen node is not the previously visited node unless this is the only neighbour node.

Let us $C_m(t)$ denote *coverage* or the fraction of the network nodes visited by any of the m random walkers at time t. $C_m(t)$ is an increasing function of t taking values between $\frac{1}{N}$ (i.e., the case when only one node is visited) and 1 (i.e., all nodes are visited).

Let us, also, define the *termination time*, denoted by T_m, as the smallest value of t such that $C_m(t) = 1$. Alternatively, it is frequently convenient to consider the *asymptotic termination time* T'_m which is defined as the smallest value of t such that $\lim_{N\to\infty} C_m(t) = 1$.

Theorem 1. *In a fully connected network topology of N nodes and m random walkers, coverage $C_m(t)$ as a function of time t is given by:*

$$C_m(t) = 1 - e^{-\frac{m}{N}t}. \tag{3}$$

3.2 A Replication Mechanism

In topologies less dense than fully connected ones, it is expected that random walkers originating from a common initiator node to frequently revisit network nodes not only due to the probabilistic nature of the random walk mechanism (as it is the case for a fully connected topology), but also due to the topology characteristics. In such a network it is expected m random walkers to cover an almost overlapping network area (frequent revisits) at the beginning, before moving to distant (and likely not previously visited) areas. Therefore, instead of m distinct movements corresponding to the m random walkers, a macroscopic observer (most likely) would observe a number of distinct movements less than m, increasing (on average) as time increases.

In order to exploit this observation and proceed with a qualitative analysis, let us assume that the underlying topology is a fully connected network (as before), in which random walkers move (and overlap) as it would have been the case if the underlying topology was not a fully connected one. The fully connected topology assumption is useful in order to simplify the analysis reusing results derived when proving Theorem 1. Let us $f(t)$ denote the (average) fraction of random walkers seen by the macroscopic observer at time t. $mf(t)$ corresponds to the (average) number of distinct movements of random walkers in the network. In general, $f(0)$ is expected to be rather small and $f(t)$ to be close to 1 for large values of t. Let us assume that $f(t) = 1 - e^{-at}$, where α is a constant that varies depending on the characteristics of each environment (e.g., number of nodes, density, bottleneck links).

Theorem 2. *In a fully connected network topology of N nodes and $mf(t)$ random walkers, coverage as a function of time t is given by:*

$$C_m(t) = 1 - e^{-\frac{m}{N}(t - \frac{1}{\alpha}(1 - e^{-at}))}. \tag{4}$$

It is interesting to observe that $C_m(t)$ increases as time increases (as expected) but not that quickly. In particular, for small values of t, $C_m(t)$ increases slowly, then it reaches an inflection point at some point $t = t_0$.

The existence of the inflection point (confirmed by simulation results), is the basic motivation behind the introduction of a simple replication mechanism in the sequel. It is evident that due to revisits, a large number of random walkers may not always allow for significant coverage improvement, while at the same time an increased number of network resources are wasted. Under replication, a small number of $m_o << m$ random walkers is initially released at the initiator node and afterwards, more random walkers are created by replicating the existing ones. Special care is taken fir the total number of random walkers in the network not to exceed m. Note that the values of m_0 comparable to m eventually do not make any difference with respect to the problem of revisits since they reduce the replication mechanism to the multiple random walkers mechanism.

The replication mechanism: Having started with m_0 random walkers, for each random walker a replica is created with constant probability $\frac{1}{q}$ after each movement. All random walkers move in the network according to the multiple random walkers mechanism.

3.3 Results

The simulation results have verified the performance covered by the analysis for a fully connected topology. For the sparser topologies, a random geometric graph was considered and simulations for various values of the variable r_c (i.e., variable that defines the connectivity of the graph) took place verifying the expected inflection point, in coverage performance. Finally, in order to examine the replication performance, a simple replication policy (consisted of one initial walker but very frequent replications) was considered and the results showed that replication outperforms the multiple random walkers mechanism when the number of simultaneous walkers (i.e., m) is large.

4 Randomly Replicated Random Walks

Taking under consideration the overall good performance of the introduced replication mechanism, when it is compared to the multiple random walkers example, a further study of its efficiency in advertising the disseminated information in various networking topologies has taken place.

The efficiency of an advertising mechanism is measured not only by the number of nodes that are informed about the location of a specific service (e.g., size of the advertising network), but also by the succeeded dissemination of this information over broader areas in the network, bringing it close to as many nodes as possible so as to reduce the intensity of their search for it (i.e., searching will apply an L-controlled flooding mechanism). As a measure of the performance regarding *stretching* the information dissemination in wider areas of the network, the previously introduced metric of L-coverage is considered. In this study, the case for $L = 0$ (i.e., to measure the size of the advertising network), $L = 1, 2$ (i.e., to measure the stretching capabilities) is examined.

To this end, the performance of two widely used techniques (flooding and single random walker) is examined through simulations, along with the performance of a proposed broad class of information dissemination schemes. The introduced Randomly Replicated Random Walks (RRRWs) scheme employs random walkers that *replicate* themselves; the *replication* policy considered here creates replicas according to an exponentially decreasing probability (in contrast with the replication policy that is assumed when comparison with the multiple walkers case took place earlier on), thus creating more replicas at the beginning of the process, controls the number of walkers and increases the probability of *stretching* the information to undiscovered parts of the underlying network. When the first replication probability equals one, the number of the walkers that are used for advertising the location of the service is large and the approach closely resembles to flooding. When the first replication probability equals zero, then only one walker is used and the approach closely resembles to SRW.

Extensive simulation results over several widely employed network topologies reveal that the RRRW scheme outperforms the single random walker (although the difference is small in the random geometric graphs), while the comparison to flooding depends on the topology. The RRRW scheme outperforms flooding in the random geometric graphs and in the clustered topologies, with respect to the size of the generated advertising network, while they manage to stretch more the information dissemination in the power-law topologies and the aforementioned clustered and random geometric graph environments, even though the generated advertising networks in the former topology are smaller in size than the ones generated by flooding. From the above, it can be stated that the RRRW scheme performs better in topologies that manage to capture best the sense of geographical coverage of a network (e.g., random geometric and clustered environments).

5 Conclusions

The study of several information dissemination techniques has taken place in this thesis. The reason for this study is an effort to examine more efficiently (both regarding the number of generated messages and the termination time) the process of Service Discovery in modern unstructured network environments.

For probabilistic flooding, analytic asymptotic bounds were given for the forwarding probability p_f in order to achieve full coverage of an underlying connected random graph ($G(N,p)$). On top of it, an analysis comparing the overhead induced both by probabilistic and by full flooding was conducted, for various coverage performance of the underlying graph $G(N,p)$. The simulation results were in accordance with the analysis.

For multiple random walkers, analytical expressions for coverage and termination time when a fully connected topology is presented. The study of less dense topologies revealed that the coverage performance in the very early stages is not so effective due to the large number of walkers that is collected near the initiator node. Later this observation was confirmed by simulation results.

A replication mechanism was introduced, mainly as a solution to the aforementioned ineffective coverage performance. This mechanism is examined as an effective advertising approach, studying whether it can stretch the dissemination of information in broader networking areas and fill the performance gap between the full flooding and single random walker approach.

References

1. A. Segall, “Distributed network protocols”, IEEE Trans. In- form. Theory, vol. IT-29, Jan. 1983.
2. Gnutella RFC, http://rfc-gnutella.sourceforge.net/, 2002.
3. C. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, “Search and Replication in Unstructured Peer-to-Peer Networks”, ICS 2002, 2002.
4. C. Gkantsidis, M. Mihail and A. Saberi, “Hybrid Search Schemes for Unstructured Peer-to-Peer Networks”, IEEE Infocom 2005, 2005.
5. N. B. Chang and M. Liu, “Optimal Controlled Flooding Search in a Large Wireless Network”, Third International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt’05), 2005, pp. 229 - 237.
6. V. Kalogeraki, D. Gunopoulos and D. Zeinalipour-Yazti, “A Local Search Mechanism for Peer-to-Peer Networks”, in CIKM (International Conference on Information and Knowledge Management), 2002.
7. F. Banaei-Kashani and C. Shahabi, “Criticality-based Analysis and Design of Unstructured Peer-to-Peer Network as “Complex Systems”, in Proceedings of the Third International Symposium on Cluster Computing and the Grid, 2003, pp. 51 - 358.
8. D. Tsoumakos and N. Roussopoulos, “Adaptive Probabilistic Search for Peer-to-Peer Networks,” 3rd IEEE International Conference on P2P Computing, 2003.
9. B. Bollobás, ”Random Graphs”, Cambridge University Press, Second Edition, 1999.
10. N. Alon, C. Avin, M. Koucky, G. Kozma, Z. Lotker and M. R. Tuttle. ”Many random walkers are faster than one”, in SPAA ’08: Proceedings of the twentieth annual symposium on Parallelism in algorithms and architectures, pages 119-128, New York, NY, USA, 2008, ACM.