Adapting Search Theory to Networks

by

Dr. A. J. Douglas

Operational Research Team
Communications Security Establishment[1]

Abstract

Mobile software objects will become much more common in the near future. For some purposes in information operations, it will be necessry to locate them within large networks. There is a large body of search theory that has examined analogous problems in other settings. This paper reports on a preliminary examination of the extent to which these results can be adapted to the network setting.

---

[1]This team is composed of Defence Scientists assigned from the Operational Research Division, National Defence Headquarters, Department of National Defence.

INTRODUCTION

General Discussion of the Problem

Information operations (IO) has recently emerged as an area of concern. Whether one traces its origins to the earliest thoughts on warfare, such as the classical writings of Sun Tzu, or focuses on the most recent developments in the technology of the networked world, certain enduring problems remain. For defensive IO, these include indications and warning of attack, and recognition and characterization of attack. In warfare, these problems are usually addressed by gathering and analysing information about what is going on, and where it is happening. For defensive information operations, the same process can be expected to apply.

This paper addresses the question "Where?" It reports on a preliminary investigation of the problem of locating activity in networked structures. This problem is addressed, in other settings, in the field of search theory. This comprises a large body of work, both theoretical and practical, developed over the last six decades. (See References 1 and 2 and the bibliographies therein.) Much of this work, such as searching for submarines or searching for presumably crashed aircraft, dealt with problems that were essentially searches on a two-dlimensional surface. The objective of the investigation reported in this paper was to determine how readily results from this existing body of knowledge could be applied in the topologically different setting of networks. Some of the results are therefore necessarily somewhat mathematical. Others, however, that address some possible operational analogies, are discursive.

Mobile Software Objects

The concept of transmiting code from one computer to another for execution there has been around for more than 40 years. It was present, for example, in the architecture of the CDC 1604 with its accompanying satellite computers, and that system was introduced in 1958 (Reference 3). The idea may well have existed earlier. As a programming practice, it has evolved from an occasional concern of the system programmers working at the deepest levels of the operating system to a common and widespread technique. It is now utterly unremarkable for a web browser to bring in a webpage containing a JAVA applet that then executes.

The next stage, in which objects move under conditions of their own determination, is already upon us. Development environments are being created to facilitate this. One of these is for the creation and use of "Aglets"--not the hardened tips of your shoelaces, but "agile applets". A description of these, including a discussion of some of the security issues they raise, can be found in Reference 4 and in the links from that webpage. The possible uses of such mobile objects appear to be limited only by the imaginations of their creators. It is likely that the networked world will see many of these.

It would be prudent to expect that a few of these uses will not be benign. Defensive IO will then have to deal with mobile attackers. It is likely that at least some defensive measures would be improved by knowledge of where the attackers are.

Search Theory

Much of search theory deals with a situation in which there is a target, a searcher, and a set of cells that each may occupy. From each cell, there are others, called adjacent cells, to which a move is possible in a single step. The searcher and the target move in turn. Moves are governed by probabilities; given the initial cell for a move, it is assumed that, for each adjacent cell, the probability of moving there is specified. With this information, one can start from specified initial positions of the searcher and target and calculate the probability that, after a specified number of steps, they occupy the same cell. For some problems, such an encounter would be called a detection; for others, detection would then occur with a specified probability. The variety of search problems arises from the differing adjacency relationships between cells, the differing transition probabilities, the differing detection probabilities, and the differing initial conditions. This is readily adapted to network problems. The cells are the different sites in the network, and adjacency relationships exist where there are direct communication links between sites.

The problem of detecting a target, given that there is an encounter with the searcher in the same cell, is going to be of increasing importance in dealing with mobile software objects. Techniques have been proposed to conceal cryptographically the capabilities of an object. One such proposal is described in Reference 5. If these or similar techniques become widely used, they will greatly complicate the problem of converting an encounter into a detection.

Among the many ways that searcher and target can move, one that has been studied in depth is the case where both move at random, choosing among the available next steps with equal probability. Besides conferring some mathematical simplifications, this kind of search is easy to implement, and thus gives a low-cost alternative in cost/benefit comparisons with other approaches. It also has an important game-theoretic advantage--it cannot be out-guessed. If you don't know what you are going to do next, your opponent isn't likely to figure it out either. This, too, makes random search an important alternative for comparison with other methods. Results for random search strategies will be discussed in the next section with the intent of using them as a basis for such comparisons and as an indication of the difficulty of the problem.

SOME MATHEMATICAL RESULTS CONCERNING RANDOM SEARCH

A Classical Result of Pólya

In Reference 6, Pólya considered networks in n-dimensional space, with the nodes being all points with integer co-ordinates and the links being lines joining them and parallel to the axes. The one-dimensional case is the real line with nodes at the integers and links consisting of the segments between them. The two-dimensional case is a square grid on the plane. He considers the case in which the target and the searcher move at random; the search is of indefinite duration. The main result is that, for the one- and two-dimensional cases, the searcher will encounter the target with probability one; for higher dimensions, the probability is less than one.

In Reference 7, the probability of the searcher encountering the target in a three-dimensional search was calculated to be about 0.35; the methods of Pólya's paper show that it will be considerably smaller in higher dimensions.

These results require some careful interpretation. First of all, Pólya's proof is heavily dependent upon the precise form of network that he studied. Adapting it to more general situations is an intriguing mathematical problem that does not appear to be trivial.

Secondly, these results apply to searches in networks of infinite extent and of indefinite duration. In any finite connected network, in which target and searcher move at random, an encounter is inevitable if the search is continued without time limit. However, if the network is extremely large, the theorem indicates that, in higher dimensions the search should be expected to be extremely protracted.

Effects of Dimension

Pólya's result is stated in terms of the dimension of the space in which the network is found. However, one of the major results of classical dimension theory establishes that a network made up of nodes that can be considered as points and links that can be considered as (possibly curved) lines can be embedded in three-dimensional space[2]. Pólya's use of higher-dimensional spaces provides a convenient way of describing networks, but this imbedding theorem shows that his result is really about the number of links leaving each node.

Pólya's result is an indication that, as the number of links per node increases, the chances of success through random search will decrease rapidly.

PRACTICAL CONSEQUENCES

Cautionary Remark

As has been indicated in several places, the mathematical theory does not exactly fit the problem. The results are therefore indicative, but not logically conclusive.

Network Layers

Networks are organized in layers, with different operations being performed in different layers. In the widely used ISO standard for Open Systems Interconnection, there are seven layers, starting at the physical layer and rising to the applications layer. In general, as one works up the layered structure, the network becomes more densely connected. This would indicate that searching might be expected to be more effective if conducted in lower network layers.

Usefulness of Random Search

---

[2]The exact statement of the theorem is: Every separable metric space of dimension less than or equal to n can be topologically imbedded in E_{2n+1}. See Reference 3, Chapter V.

One should expect random search to find targets efficiently only in cases where the networks are lightly connected, and then only in cases where there is good reason to believe that the search will start close to the target.

If the network is densely connected, or if the initial whereabouts of the target is very imprecise, random search can be expected to be protracted and in most cases unsuccessful.

Non-Random Searching

Non-random searching can be systematic, in that it attempts to cover all or part of the network exhaustively, or it can be adaptive, in that it attempts to learn something about the behavior of the target and adjust its actions according. If the target is not just mobile but also evasive, these approaches lead to the problems of game theory. There is an extensive literature on these, but reviewing it is beyond the scope of this paper.

Unless there is information limiting the location of a target to a tractable portion of the network, exhaustive searches are unlikely to be useful for long. In particular, when the Internet adopts addresses consisting of four groups of 32 bits, as is now planned, exhaustive searching becomes infeasible.

Adaptive methods are inherently empirical. Although theory can often be of some help, it is essential to gather extensive data on the behavior of targets of interest and from this data to deduce their patterns of movement.

Tracking

One should expect to have to deal with richly linked networks in which targets are very imprecisely located. It may be necessary to develop software agents that can follow targets once they have been found. This could be done both as a means of dealing with that particular target and also to gather data on target behavior in support of adaptive searching.

SUMMARY

The following is a summary of the points in this paper:

1. Mobile software objects exist now, and will become more common in the future; one can confidently expect some of these to be harmful.

2. Some networks will become large enough that exhaustive searching will be infeasible, and richly connected enough that random searching, even about a known starting point, will usually be unsuccessful.

3. There will then be two possibilities:

1. Tracking objects of interest as they move; and,

2. Adaptive searching using knowledge of the behavior of targets of interest.

4. These possibilities both require the development of software agents, in the first case to do the tracking and in the second case to gather data on the behavior of objects of interest; agents for these purposes could be quite similar.

REFERENCES

1. Koopman, Bernard O., Search and Screening: General Principles with Historical Applications; Pergammon Press, New York, 1980.

2. Stone, Lawrence D., Theory of Optimal Search; Academic Press, New York, 1975.

3. **Lee, J.A.N. and Stanley Winkler, with copy by Merlin Smith, Key Events in the History of Computing; prepared for the IEEE Computer Society for distribution in 1996 as part of the 50th Anniversary Activities; available on the Internet at http://ei.cs.vt.edu/~history/50th/30.minute.show.html**

4. **Aglets Software Development Kit; IBM Corporation, 1996; available on the Internet at http://www.trl.ibm.co.jp/aglets/documentation.html**

5. **Riordan, J.** and B. Schneier, Environmental Key Generation towards Clueless Agents; Mobile Agents and Security, G. Vigna, ed., Springer-Verlag, 1998, pp. 15-24; available on the Internet at http://www.counterpane.com/clueless-agents.html

6. Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz, by Georg Pólya; Mathematische Annalen, vol. 84 (1921), pp. 149-160.

7. Random paths in two and three dimensions, by W. H. McCrea and F. J. W. Whipple; Proceedings of the Royal Society of Edinburgh, vol. 60 (1940) pp. 281-298.

8. Dimension Theory, by Witold Hurewicz and Henry Wallman; Princeton University Press, Princeton, New Jersey, 1948.