Study and Analysis of Ant System

By Pawandeep Chahal

Desh Bhagat Institute of Engg. & Management, Moga

Abstract – A lot of species of ants have a trail-laying/trailfollowing behavior when foraging. While moving, individual ants deposit on the ground a volatile chemical substance called pheromone, forming in this way pheromone trails. Ants can smell pheromone and, when choosing their way, they tend to choose, in probability, the paths marked by stronger pheromone concentrations. In this way they create a sort of attractive potential field, the pheromone trails allows the ants to find their way back to food sources (or to the nest). Also, they can be used by other ants to find the location of the food sources discovered by their nest mates.

GJCST-E Classification: *I.6.4*

Study and Analysis of Ant System

Strictly as per the compliance and regulations of:
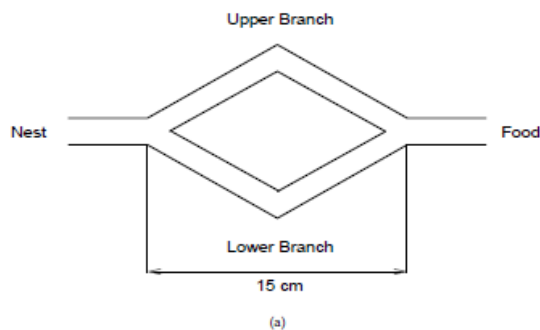
Study and Analysis of Ant System

Pawandeep Chahal

I. Ant System

A lot of species of ants have a trail-laying/trail-following behavior when foraging. While moving, individual ants deposit on the ground a volatile chemical substance called pheromone, forming in this way pheromone trails. Ants can smell pheromone and, when choosing their way, they tend to choose, in probability, the paths marked by stronger pheromone concentrations. In this way they create a sort of attractive potential field, the pheromone trails allows the ants to find their way back to food sources (or to the nest). Also, they can be used by other ants to find the location of the food sources discovered by their nest mates.

a) Binary bridge experiment with same branch length

Let us consider first the binary bridge experiment [20] whose setup is shown in Figure 1 (a). The nest of a colony of Argentine ants Linepithema humile and a food source have been separated by a diamond-shaped double bridge in which each branch has the same length. Ants are then left free to move between the nest and the food source. The percentage of ants which choose one or the other of the two branches is observed over time. The result in Figure 1 (b) is that after an initial transitory phase lasting few minutes during which some oscillations can appear, ants tend to converge on a same path.

In this experiment initially there is no pheromone on the two branches, which are therefore selected by the ants with the same probability. Nevertheless, after an initial temporary phase, random fluctuations cause a few more ants to randomly select one branch, the upper one in the experiment shown in Figure 1 (a). Since ants deposit pheromone while walking back and forth the greater number of ants on the upper branch determines a greater amount of pheromone on it, which in turn stimulates more ants to choose it, and so on in a circular way.
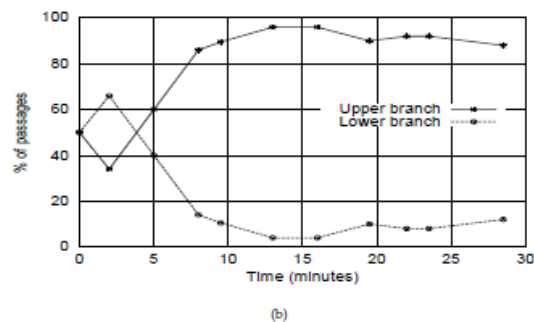

Figure 1 : Binary bridge experiment with same branch length. (a) Experimental setup. (b) Results for a typical single trial [20]

To describe this convergent behavior of the ants, the experiment has proposed a probabilistic model which closely matches the experimental observations. It is assumed that the amount of pheromone on a branch is proportional to the number of ants which have been using the branch in the past. This assumption implies that the pheromone trail is persistent, that is, pheromone trail does not evaporate. Given that an experiment typically lasts approximately one hour, it is plausible to assume that the amount of pheromone evaporated in this time period is negligible. For longer durations, pheromone evaporation must be taken into account. In the model, the probability of choosing a branch at a certain time depends on the total amount of pheromone on the branch, which, in turn, is proportional to the number of ants which have used the branch until that moment. More precisely, let U_m and L_m be the numbers of ants which have used the upper and lower branch after a total of m ants have crossed the bridge, $U_m + L_m$ = m. The probability $P_U(m)$ with which the (m + 1)th ant chooses the upper branch is

$$P_U(m) = \frac{(U_m+k)^h}{(U_m+k)^h+(L_m+k)^h} \qquad (2.1)$$

While the probability $P_L(m)$ that the ant chooses the lower branch is

Author : Research Scholar, CMJ University, Shillong , Meghalaya, Asst. Prof, Department of IT, Desh Bhagat Institute of Engg. & Management, Moga.

$$P_L(m) = 1 - P_U(m) \qquad (2.2)$$

This functional form for the probability of choosing a branch over the other was obtained from experiments on trail following [62]; the parameters h and k allow to fit the model to experimental data. The dynamics regulating the ant choices follows from the above equation:

$$\begin{cases} U_m + 1 = U_m + 1, & if\ \psi \le P_U \\ U_m + 1 = U_m, & otherwise \end{cases} \qquad (2.3)$$

Where ψ is a random variable uniformly distributed over the interval [0,1]. Monte Carlo simulations were run to test the correspondence between this model and the real data: results of simulations were in agreement with the experiments with real ants when parameters were set to $k \approx 20$ and $h \approx 2$ [62].

b) *Binary bridge experiment with different branch length*

The previous experiment shows how the presence of pheromone affects in general the ant decisions and constrains the foraging behavior of the colony as a whole. If the branches of the bridges are of different length, then the pheromone field can lead the majority of the ants in the colony to select the shortest between the two available paths. In this case, the first ants able to arrive at the food source are those that traveled following the shortest branch (as in Figure 2.2). Accordingly, the pheromone that these same ants have laid on the shortest branch while moving forward towards the food source makes this branch marked by more pheromone than the longest one. The higher levels of pheromone present on the shortest branch stimulate these same ants to probabilistically choose again the shortest branch when moving backward to their nest. This recursive behavior can be thoroughly described as an autocatalytic effect because the very fact of choosing a path increases its probability of being chosen again in the near future.

During the backward journey, additional pheromone is released on the shortest path. In this way, pheromone is laid on the shortest branch at a higher rate than on the longest branch. This reinforcement of the pheromone intensity on the shorter paths is the result of a form of implicit path evaluation: the shorter paths are completed earlier than the longer ones, and therefore they receive pheromone reinforcement more quickly. Therefore, for a same number of ants choosing either the shortest or the longest branch at the beginning, since the pheromone on the shortest branch is accumulated at a higher rate than on the longest one, the choice of the shortest branch becomes more and more attractive for the subsequent ants at both the decision points. The experimental observation is that, after a transitory phase which can last a few minutes, most of the ants use the shortest branch. It is also observed that the colony's probability of selecting the shortest path increases with the difference in length between the long and the short branches.

(a)   (b)   (c)
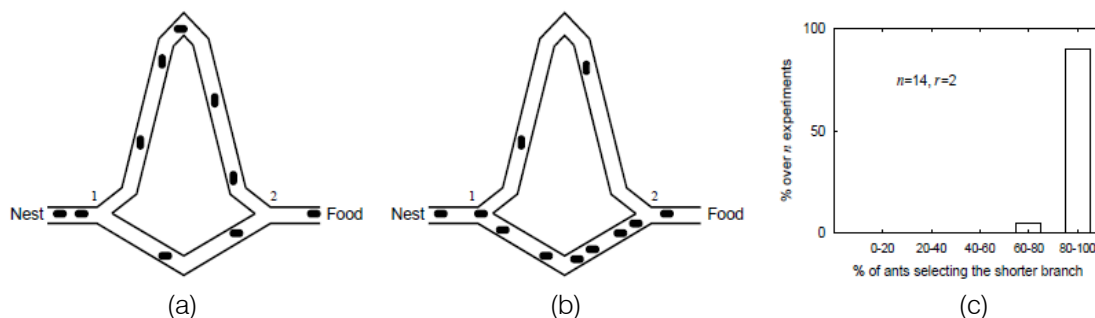
Figure 2 : Experiment with different branch length. (a) Ants start exploring the bridge. (b) Eventually most of the ants choose the shortest path. (c) Distribution of the percentage of ants that selected the shorter branch

In this experiment, the importance of initial random fluctuations is much reduced with respect to the previous experiment. In Figure2 (a) – (b) are shown, together with the experimental apparatus, the typical result of an experiment with a bridge with branches of different lengths. Figure 2 (c) shows the distribution of the results over n=14 experiments for the case of a bridge in which the length r=2 of the longer branch is twice that of the shorter one.

Figure 2.3 shows in a schematic way how the effect of round-trip pheromone laying/sensing can easily determine the convergence of all the ants on the shortest between two available paths. At time t = 0 two ants leave the nest looking for food. According to the fact that no pheromone is present on the terrain at the nest site, the ants select randomly the path to follow. One ant chooses the longest and one the shortest path bringing to the food. After one time unit, the ant that chose the shortest path arrives at the food reservoir. The other ant is still on its way. The intensity levels of the pheromone deposited on the terrain are shown; where the intensity scale on the right says that a darker color means more pheromone. Pheromone evaporation is considered as negligible according to the time duration of the experiment.

Figure 2.3 : Example of how the effect of laying/sensing pheromone during the forward and back journeys from the nest to food sources can easily determine the convergence of all the ants of the colony to the shortest between two available paths

The ant already arrived at the food site must select the way to go back to the nest. According to the intensity levels of the pheromone near the food site, the ant decides to go back by moving along the same path, but in the opposite direction. Additional pheromone is therefore deposited on the shortest branch. At t = 2 the ant is back to the nest, while the other ant is still moving toward the food along the longest path. At t = 3 another ant moves from the nest looking for food. Again, he/she selects the path according to the pheromone levels and, therefore, it is biased toward the choice of the shortest path. It is easy to imagine how the process iterates, bringing, in the end, the majority of the ants on the shortest path.

II. ROBUSTNESS AND ADAPTIVITY

Even if it is not always true that the shortest path behavior will arise, it is often the case that alternative non-random, self-organized, global patterns of activity will arise. That is, under reasonable conditions (e.g., environmental conditions are not such that pheromone evaporates faster than the average time necessary for an ant to reach the target), some interesting regular patterns can be eventually observed. This fact witnesses the overall robustness of the mechanisms at work in ant colonies, as well as the fact that they are able to produce an interesting variety of different organized behaviors. These are key properties in real-world environments, which require robustness, adaptivity and the ability to provide satisfactory responses to a range of possible different situations.

The general robust collective behavior of ant colonies with respect to variations in the values of the external conditions is a key-aspect of their biological success. They, like other classes of social insects, are crystalline examples of natural complex adaptive systems that the evolutionary pressure has made sufficiently robust to a wide range of external variations.

III. CONNECTION-ORIENTED AND CONNECTIONLESS PROTOCOLS

Protocols can be either connection-oriented or connectionless in nature. In connection-oriented protocols, corresponding entities maintain state information about the dialogue they are engaged in. This connection state information supports error, sequence and flow control between the corresponding entities. The windowing scheme presented earlier is an example of a connection-oriented protocol.

Error control refers to a combination of error detection (and correction) and acknowledgment sufficient to compensate for any unreliability inherent to the channel. Sequence control refers to the ability for each entity to reconstruct a received series of messages in the proper order in which they were intended to be received; this is essential to being able to transmit large files across dynamically-routed mesh networks. Flow control refers to the ability for both parties in a dialogue to avoid overrunning their peer with too many messages. Connection-oriented protocols operate in three phases. The first phase is the connection setup phase, during which the corresponding entities establish the connection and negotiate the parameters defining the connection. The second phase is the data transfer phase, during which the corresponding entities exchange messages under the auspices of the

connection. Finally, the connection release phase is when the correspondents "tear down" the connection because it is no longer needed.

Networks may be divided into different types and categories according to four different criteria:

a) *Geographic spread of nodes and hosts*

When the physical distance between the hosts is within a few kilometers, the network is said to be a **Local Area Network** (LAN). LANs are typically used to connect a set of hosts within the same or a set of closely-located buildings). For larger distances, the network is said to be a **Metropolitan Area Network** (MAN) or a **Wide Area Network** (WAN). MANs cover distances of up to a few hundred kilometers and are used form interconnecting hosts spread across a city. WANs are used to connect hosts spread across a country, a continent, or the globe. LANs, MANs, and WANs usually coexist: closely-located hosts are connected by LANs which can access hosts in other remote LANs via MANs and WANs.

b) *Access restrictions*

Most networks are for the private use of the organizations to which they belong; these are called **private networks**. Networks maintained by banks, insurance companies, airlines, hospitals, and most other businesses are of this nature. **Public networks**, on the other hand, are generally accessible to the average user, but may require registration and payment of connection fees. Internet is the most-widely known example of a public network. Technically, both private and public networks may be of LAN, MAN, or WAN type, although public networks, by their size and nature, tend to WANs.

c) *Communication model employed by the nodes*

The communication between the nodes is either based on a **point-to-point** model or a **broadcast** model. In the point-to-point model, a message follows a specific route across the network in order to get from one node to another. In the broadcast model, on the other hand, all nodes share the same communication medium and, as a result, a message transmitted by any node can be received by all other nodes. A part of the message (an address) indicates for which node the message is intended. All nodes look at this address and ignore the message if it does not match their own address.

d) *Switching model employed by the nodes*

In the point-to-point model, nodes either employ **circuit switching** or **packet switching**. Suppose that a host A wishes to communicate with another host B. In circuit switching, a dedicated communication path is allocated between A and B, via a set of intermediate nodes. The data is sent along the path as a continuous stream of bits. This path is maintained for the duration of communication between A and B, and is then released.

In packet switching, data is divided into packets (chunks of specific length and characteristics) which are sent from A to B via intermediate nodes. Each intermediate node temporarily stores the packet and waits for the receiving node to become available to receive it. Because data is sent in packets, it is not necessary to reserve a path across the network for the duration of communication between A and B. Different packets can be routed differently in order to spread the load between the nodes and improve performance. However, this requires packets to carry additional addressing information.

IV. DATA COMMUNICATION AND NETWORKING

The major criteria that a Data Communication Network must meet are:
i.   Performance
ii.   Consistency
iii.   Reliability,
iv.   Recovery and
v.   Security

V. PERFORMANCE

Performance is the defined as the rate of transferring error free data. It is measured by the Response Time. Response Time is the elapsed time between the end of an inquiry and the beginning of a response. Request a file transfer and start the file transfer. Factors that affect Response Time are:

a. Number of Users: More users on a network - slower the network will run

b. Transmission Speed: speed that data will be transmitted measured in bits per second (bps)

c. Media Type: Type of physical connection used to connect nodes together

d. Hardware Type: Slow computers such as XT or fast such as Pentiums

e. Software Program: How well is the network operating system (NOS) written

VI. CONSISTENCY

Consistency is the predictability of response time and accuracy of data.

a. Users prefer to have consistent response times, they develop a feel for normal operating conditions. For example: if the "normal" response time is 3 sec. for printing to a Network Printer and a response time of over 30 sec happens, we know that there is a problem in the system!

b. Accuracy of Data determines if the network is reliable! If a system loses data, then the users

18

will not have confidence in the information and will often not use the system.

VII. RELIABILITY

Reliability is the measure of how often a network is useable. MTBF (Mean Time Between Failures) is a measure of the average time a component is expected to operate between failures. Normally provided by the manufacturer. A network failure can be: hardware, data carrying medium and Network Operating System.

VIII. RECOVERY

Recovery is the Network's ability to return to a prescribed level of operation after a network failure. This level is where the amount of lost data is nonexistent or at a minimum. Recovery is based on having Back-up Files.

IX. SECURITY

Security is the protection of Hardware, Software and Data from unauthorized access. Restricted physical access to computers, password protection, limiting user privileges and data encryption are common security methods. Anti-Virus monitoring programs to defend against computer viruses are a security measure.

X. NETWORK HIERARCHY

A general world-wide communication network consists of three parts as an access network that offers connectivity to residential users, an edge network that combines several access networks (and possibly corporate networks) and a core network (or the backbone). Important access networks are the residential telephony network, cable TV network, ADSL network and the mobile networks as GSM and Wireless LANs. As core networks, we mention the international telephony trunks and the Internet backbone(s). Edge networks lie in between and are not clearly defined but only relatively with respect to the access and the core network. The large differences in throughput and other quantities demonstrate the need for different network management and underlying physical transmission technologies.

XI. TYPES OF COMMUNICATION INTERACTION

In networking, various types of interactions between communicating parties exist.

XII. CENTRALIZED VERSUS DISTRIBUTED CONTROL

In a centralized approach, a master is appointed among the systems in a network. The master controls each interaction in the network of these systems. The other alternative, a distributed control consists in using a policy or protocol of communication (e.g. start speaking as soon as someone else stops, but back-off immediately if a third one starts). It is the rule of the protocol that controls distributed interaction.

XIII. FINITE STATE MACHINE INTERACTION

A first type of interaction between the set of systems is that driven by a finite state machine. A finite state machine follows and executes rules or actions depending on state of the process. For example, system 2 has just spoken, thus, we move to and poll system 3, and so on. The operation of a finite state machine can be visualized and described by a graph that relates the processes in the different states. A well known example of such a graph is a Petri net.

a) Client-server interaction: "ask-when-needed" or event driven

Another mode of interaction only operates when an inner state or a process in a system requires information from other connected systems. This mode is event driven and called a client-server interaction. For example, when clicking on a link at a webpage, there is a short communication with a server that returns the IP address of the machine on which the content of the intended page is stored. A client-server interaction is thus a relation between processes in which each process can take the initiative to communicate with another process. A particular process is not necessary always client or server, but a process is a client or a server with respect to another process and their role can change over time. Also a process A can be client in the relation with process B, but it can be the server with respect to process C.

The communication in distributed systems needs to be designed to avoid a "deadlock" that is the situation in which processes are waiting infinitely long for each other. The communication relations between all processes in a distributed system can be represented by a graph. Dead-locks can be avoided if that graph is a-cyclic, i.e. the graph does not contain cycles. In an a-cyclic graph or tree, there is only 1 path between each pair of processes. Since a client-server interaction asks a question and waits for the reply (using a single path in the process relation graph), the client-server concept allows to build a dead-lock free architecture of a distributed system. Large and complex distributed systems can be built based on the relatively simple client-server principle.

b) Summary of interaction models

Client-server interaction forms the basis of most network communications and is fundamental because it helps us understand the foundation on which distributed algorithms are built. Usually, there are more clients than servers. Typically, a finite state machine is used when there is little interaction or the interaction is simple such that the number of possible states in the communication protocol is limited.

Client-server interaction is more flexible and suited for intense interaction. Although we may argue that distributed networking is preferable in terms of scalability and robustness the back side of the medal is that distributed networking is more difficult to control and design. For example, distributed routing seems very robust and scalable. In nature, ants find their way individually by using a small set of rules. It seems interesting to transfer the rules deduced from the behavior of ants to communication networking. Apart from finding the minimal set of rules, the demonstration that the ensemble of rules operates correctly for the whole colony of packets in the network is difficult. Further, proving optimality or how close distributed networking lies to a global optimum is generally difficult.

XIV. COMMUNICATION MODES

In general, four different communication modes can be distinguished: unicast, multicast, broadcast and any cast. Unicast is a communication between two parties (one-to-one) and a typical example is a telephone call. Multicast consists of the modes one-to-many and many-to-many with as example a video-conference. Broadcast defined as a communication from one user to all users in a network is an extreme case of multicast. The typical example is the broadcasting of information for television and radio. Finally, anycast is a communication from one-to-any of a group. For example, when information is replicated over many servers, a user wants to download the information from an arbitrary server of that group. Most often, the anycast mode will point or route the user's request to that server nearest to the user. The user does not need to know the location or the individual addresses of the servers, only the anycast address of the group of servers.

XV. PERFORMANCE METRICS

In the design of communications networks, the preference of algorithm or implementation A of a network functionality above algorithm B depends on various factors. Beside the monetary cost, the most common technical factors, called performance metrics, are the computation complexity, the throughput, the blocking, the reliability, the security, the memory consumption, and the manageability. In general, the performance metrics for a particular algorithm/implementation are not always easy to compute. The precise definitions, the analysis, and the computation of performance metrics belong to the domain of performance analysis for which we refer to Van Mieghem (2006). Apart from the precise evaluation of the performance of a particular algorithm/ implementation, some of the performance metrics are not yet universally and accurately defined. For example, the reliability or the robustness of a network topology can be evaluated in many different ways. Another frequently appearing term as a design metric is the scalability.

XVI. SCALABILITY

The term "scalability" expresses the increase in the complexity to operate, control or manage a network if relevant network parameters such as the size or the number of nodes/systems in the network, the traffic load, the interaction rate, etc. increase. Whether a property of a network is scalable or not strongly depends on that property itself.

BIBLIOGRAPHY

1. T. Abe, S. A. Levin, and M. Higashi, editors. Biodiversity: an ecological perspective. Springer-Verlag, New York, USA, 1997.
2. M. Beckmann, C.B. McGuire, and C.B. Winstein. Studies in the Economics of Transportation. Yale University Press, 1956.
3. R. Bellman. Dynamic Programming. Princeton University Press, 1957.
4. R. Bellman. On a routing problem. Quarterly of Applied Mathematics, 16(1): 87–90, 1958.
5. D. Bertsekas, J. N. Tsitsiklis, and Cynara Wu. Rollout algorithms for combinatorial optimization. Journal of Heuristics, 3(3):245–262, 1997.
6. M. Birattari, G. Di Caro, and M. Dorigo. Towards the formal foundation of ant programming. In M. Dorigo, G. Di Caro, and M. Sampels, editors, Ants Algorithms - Proceedings of ANTS 2002, Third International Workshop on Ant Algorithms, Brussels, Belgium, September 12–14, 2002, volume 2463 of Lecture Notes in Computer Science, pages 188–201. Springer-Verlag, 2002.
7. J.A. Bland. Optimal structural design by Ant Colony Optimization. Engineering optimization, 33: 425–443, 2001.
8. J.A. Bland. Layout of facilities using an Ant System approach. Engineering optimization, 32(1), 1999. [38] J.A. Bland. Space-planning by Ant Colony Optimization. International Journal of Computer Applications in Technology, 12, 1999.
9. K. Bolding, M. L. Fulgham, and L. Snyder. The case for chaotic adaptive routing. Technical Report CSE-94-02-04, Department of Computer Science, University of Washington, Seattle, 1994.
10. E. Bonabeau, M. Dorigo, and G. Theraulaz. Swarm Intelligence: From Natural to Artificial Systems. Oxford University Press, 1999.
11. J. Broch, D.A. Maltz, D.B. Johnson, Y.C. Hu, and J. Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. In Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom98), 1998.

12. D. Camara and A.F. Loureiro. Gps/ant-like routing in ad hoc networks. Telecommunication Systems, 18(1–3):85–100, 2001.

13. H.S. Chang, W. Gutjahr, J. Yang, and S. Park. An Ant System approach to Markov Decision Processes. Technical Report 2003-10, Department of Statistics and Decision Support Systems, University of Vienna, Vienna, Austria, September 2003.

14. C. Cheng, R. Riley, S.P. Kumar, and J.J. Garcia-Luna-Aceves. A loop-free extended bellman-ford routing protocol without bouncing effect. ACM Computer Communication Review (SIGCOMM '89), 18 (4):224–236, 1989.

15. L. Chrisman. Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. In Proceedings of the Tenth National Conference on Artificial Intelligence, pages 183–188, 1992.

16. I. Cidon, R. Rom, and Y. Shavitt. Multi-path routing combined with resource reservation. In IEEE INFOCOM'97, pages 92–100, 1997.

17. I. Cidon, R. Rom, and Y. Shavitt. Analysis of multi-path routing. IEEE/ACM Transactions on Networking, 7(6):885–896, 1999.

18. I. Colorni, M. Dorigo, V. Maniezzo, and M. Trubian. Ant system for job-shop scheduling. Belgian Journal of Operations Research, Statistics and Computer Science (JORBEL), 34:39–53, 1994.

19. M. Cottarelli and A. Gobbi. Estensioni dell'algoritmo formiche per il problema del commesso via ggiatore. Master 's thesis, Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy, 1997.

20. J.-L. Deneubourg, S. Aron, S. Goss, and J.-M. Pasteels. The self-organizing exploratory pattern of the argentine ant. Journal of Insect Behavior, 3:159–168, 1990.

21. J.L. Deneubourg, S. Goss, N. Franks, A. Sendova-Franks, C. Detrain, and L. Chretien. The dynamics of collective sorting: robot-like ants and ant-like robots. In S. Wilson, editor, Proceedings of the First International Conference on Simulation of Adaptive Behaviors: From Animals to Animats, pages 356–365. MIT Press, Cambridge, MA, USA, 1991.

22. G. Di Caro and M. Dorigo. AntNet: A mobile agents approach to adaptive routing. Technical Report 97–12, IRIDIA, University of Brussels, Belgium, June 1997.

23. G. Di Caro, F. Ducatelle, and L.M. Gambardella. AntHocNet: an ant-based hybrid routing algorithm for mobile ad hoc networks. In Proceedings of Parallel Problem Solving from Nature (PPSN) VIII, volume 3242 of Lecture Notes in Computer Science, pages 461–470. Springer-Verlag, 2004.

24. G. Di Caro, F. Ducatelle, and L.M. Gambardella. AntHocNet: an adaptive nature-inspired algorithm for routing in mobile ad hoc networks. European Transactions on Telecommunications, 16(5), October 2005.

25. E. W. Dijkstra. A note on two problems in connection with graphs. Numerical Mathematics, 1:269–271, 1959.

26. M. Dorigo, G. Di Caro, and L. M. Gambardella. Ant algorithms for discrete optimization. Artificial Life, 5(2):137–172, 1999.

27. J.J. Garcia-Luna-Aceves and S. Murthy. A path-finding algorithm for loop-free routing. IEEE/ACM Transactions on Networking, February 1997.

28. F. Glover. Tabu search, part I. ORSA Journal on Computing, 1:190–206, 1989.

29. J. Moy. OSPF version 2. Request For Comments (RFC) 1583, Network Working Group, 1994.

30. J. Moy. OSPF Anatomy of an Internet Routing Protocol. Addison-Wesley, 1998.

31. G. Owen. Game Theory. Academic Press, third edition, 1995.

32. M.L. Puterman. Markov Decision Problems. John Wiley & Sons, 1994.

33. Rome, Kay, and Friedemann Mattern. The design space of wireless sensor networks. IEEE Wireless Communications, 11(6):54–61, December 2004.

This page is intentionally left blank