

A bio-inspired crime simulation model

Vasco Furtado^{a,*}, Adriano Melo^a, André L.V. Coelho^a, Ronaldo Menezes^b, Ricardo Perrone^c

^a Graduate Program in Applied Informatics, University of Fortaleza (Unifor), Ceará, Brazil

^b Computer Sciences, Florida Tech, Melbourne, Florida, USA

^c Laboratory of Distributed Systems, Federal University of Bahia, Salvador-BA, Brazil

ARTICLE INFO

Article history:

Received 20 August 2008

Received in revised form 3 August 2009

Accepted 30 August 2009

Available online 8 September 2009

Keywords:

Crime simulation

Bio-inspired systems

Ant colony optimization

Genetic algorithms

Social networks

Multiagent simulation

ABSTRACT

In this paper we describe a multiagent crime simulation model that resorts to concepts of self-organizing bio-inspired systems, in particular, of the Ant Colony Optimization algorithm. As the matching between simulated and real crime data distributions depends upon the tuning of some control parameters of the simulation model (in particular, of the initial places where criminals start out), we have modeled the calibration of the simulation as an optimization problem. The solution for the allocation of criminals into gateways is also undertaken by a bio-inspired method, namely, a customized Genetic Algorithm. We show that this approach allows for the automatic discovery of gateway configurations that, when employed in the simulation, produce crime distributions that are statistically close to those observed in real data.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Recently, an extensive analysis conducted over real crime data related to a large Brazilian metropolis [11] demonstrated that the spatial distribution of crimes such as robberies, thefts, and burglaries follows a power law, more specifically, a Zipfian distribution [40]. This means that the frequency of crime occurrences related to a specific geographic area, when considered per type of crime, tends to scale according to a power-law distribution, yielding the formation of hot spots [34]. In the same work [11], an analysis over the temporal aspect reveals that these crime events follow an exponential distribution per period of analysis.

Although knowing the crime distribution profile for a given moment may be necessary to better conduct some of the police decision-making activities, it is not enough to help one gain further insights into crime in its totality. This is because crime is a dynamic phenomenon, and the decision of protecting a frequently-attacked target at a given point in time eventually leads to the exposure of other potential targets in the future, due to a range of restrictions in terms of resources availability (e.g., human resources).

In this sense, we advocate that a better understanding of the trends of criminal activities and the types of reactions criminals might potentially undertake is a crucial task to be pursued. In this context, the goal of the research we have been conducting in the last years

[18,19,31] is to produce a crime simulation system that reproduces crime phenomena as realistically as possible. Our ultimate goal is to uncover strategies for police patrolling (more precisely, police patrol routes) that could cope well with the dynamics of crime when criminal agents are capable of learning “on the fly.”

In this article, we provide a major step toward the aforementioned goal by introducing a dynamic model of crime against property¹ that shows experimentally how this type of crime evolves. The main challenge behind this effort lies in the definition of a simulation model that could generate crime episodes according to a spatial Zipfian distribution and, at the same time, be in agreement with real data. For such a purpose, we have designed a multiagent criminal model that mimics real-life criminal behavior in consonance with some sociological studies [1,36], paying special attention to the following facts: (i) the environment where the agents live is a digitalized map representing the real-life area; (ii) criminals improve their performance over time by creating preferences according to their experience in crime; and (iii) social communication among criminals must also be properly modeled, because criminal behavior depends not only on individual incentives but also on the behavior of the perpetrators' peers and neighbors.

One distinctive aspect of the conceived criminal model is that it resorts to concepts related to self-organizing bio-inspired systems, in particular the Ant Colony Optimization (ACO) algorithm [7,15]. The rationale behind this choice is twofold. Through an ant-based

* Corresponding author.

E-mail addresses: vasco@unifor.br (V. Furtado), aanmelo@gmail.com (A. Melo), acoelho@unifor.br (A.L.V. Coelho), rmenezes@cs.fit.edu (R. Menezes), perrones@ufba.br (R. Perrone).

¹ Crime against property includes burglaries, robberies, etc. as long as the main target is an establishment. Hence, situations like a *person being mugged on the street* are not included in this study.

perspective, we can model criminals as agents that account for both the individual and social aspects we intend to consider in our crime simulation model. They are endowed with the capability to pursue self-organized behavior by considering their individual (local) activities as well as the influence of other criminals in the community they live in. At the same time, as we have identified experimentally in [18], ant-based multiagent systems are capable of reproducing spatial behavior of a power-law nature. Basically, we have shown that two features of the model lead to this important result: the possibility of social communication between criminals, and the fact that they follow a kind of preferential attachment mechanism (popularly called the *rich get richer* phenomenon) [5,25]. In the context of crime, the preferential attachment mechanism reflects the criminal preference to commit crimes in places where they feel comfortable because of their past experience and learning.

Even considering these interesting capabilities of the model, we have to point out that the matching between simulated and real crime data distributions depends very much upon the tuning of some control parameters of the simulation model; in particular, the initial places where criminals start out in the simulation process (henceforth called gateways). As this particular tuning is not trivial per se, we have decided to model it as an optimization problem. Therefore, in this paper we describe a solution for the allocation of criminals into gateways via another bio-inspired technique, namely, a customized Genetic Algorithm (GA) [17]. We show that this approach allows for the automatic discovery of gateway configurations that, when employed in the simulation, produce crime distributions that are statistically close to those observed in real data. The experiments reported here compare the matching between the spatial distribution of crimes generated from our simulator and the actual crime distribution data recorded for a region of Fortaleza, a large Brazilian urban metropolis.

The remainder of the paper is structured as follows. First, we describe related work with the theoretical basis upon which we have constructed and tuned our crime simulation model. Then, we focus on the characterization of the elements behind this model, paying special attention to the description of the criminal learning behavior and to the GA-based solution for the criminal–gateway allocation problem. The experimental results come next, evidencing that the whole bio-inspired approach is indeed capable of generating adequate criminal–gateway configurations and simulated crime distributions that are closely related to those observed in real data. We conclude by describing the relevance of these findings and providing directions for future work.

2. Multiagent simulation, ant colony optimization, and genetic algorithms

Multiagent Systems (MAS) [37] involve the study of the behavior of autonomous and organized groups of agents with the purpose of providing distributed, emergent solutions to complex problems that could not be achieved by each individual agent alone. On the other hand, the deployment of simulations for the purpose of gaining insights along a given decision-making process can be a very effective approach one could resort to, as computer simulations usually allow the focused analysis of important issues by investigating their influences, either separately or conjointly.

Recently, multiagent systems have been successfully adopted in conjunction with simulation models, as the inherent characteristics of the former (e.g., agent autonomy, reactivity, and pro-activity) facilitate the construction and simulation of more realistic and dynamic models, thus contrasting directly with conventional computer simulation approaches. The outcome is generally referred to as Multiagent-based Simulation (MABS) systems, which—according to [20]—are especially appropriate when one has to deal with interdisciplinary problem domains, such as the public-safety domain inves-

tigated here. In particular, we advocate that the multiagent approach (bottom-up in nature) is appropriate for the study of social and urban problems, since social/urban environments are dynamic, non-linear, and composed of a great number of variables and entities. As pointed out by [16], some of the main goals behind the construction of MABS systems are the following:

- To test hypotheses related to the emergence of macro-level behavior from interactions occurring at micro levels;
- To build theories that can contribute to a better understanding of sociological, psychological, and ethological phenomena; and
- To integrate partial theories coming from different disciplines (e.g., sociology, cognitive psychology, and etiology) into a common theoretical framework.

The study of agent self-organization, and related concepts such as emergence, is based on the idea that societies of agents demonstrate intelligent behavior at the collective level out of simple rules pertaining to the individual level. What is interesting behind this paradigm is that the individual rules, when considered alone, cannot explain the behavior that emerges at the collective level. Within this context, Swarm Intelligence (SI) [7] has come forth as the discipline devoted to the study of biological systems characterized by (i) strictly local communication; (ii) the formation of emergent spatial–temporal structures; and (iii) stochastic decisions made by the agents based solely on the local information available. One of the most well-known branches of SI deals specifically with the study of novel optimization algorithms inspired by the social behavior exhibited by some species of ants. Arguably, the main product in this line of research is the Ant Colony Optimization (ACO) algorithm [14,15], a population-based metaheuristic that has shown promising results while tackling combinatorial optimization problems that can be represented as graphs, mainly those with dynamic settings [23].

In a nutshell, ACO works as follows. Agents (ants) are endowed with the capability to explore the discrete space of solutions related to a given problem. In doing so, they leave feedback information (normally in the form of pheromone marks) on the space itself, signaling about visited locations (i.e. building blocks) associated with satisfactory solutions. On the other hand, the path each individual ant takes is directly influenced by the pheromone marks left by their peers in the environment; so, the larger the amount of pheromone in a given location, the more attractive that location becomes for being visited by the whole population of ants. By this means, even more satisfactory solutions are able to emerge by putting together those building blocks with higher levels of pheromone. In order to avoid early convergence to local optima solutions, the approach assumes that the pheromone marks are volatile; that is, the pheromone information is short-lived and, without reinforcement activity, the “hints” left in that position start to fade with time. Although it had never been explored for the purposes of modeling criminal behavior, ACO—due to its interesting properties—seemed to us to be an ideal fit to our purposes.

Like ACO, Genetic Algorithms (GA) comprehend a prominent bio-inspired population-based metaheuristic, which, in turn, is based on the mechanics of natural selection and genetics [17]. According to the GA framework, candidate solutions (referred to as chromosomes or individuals) to a given continuous/discrete optimization problem play the role of individuals in a population, while the cost (fitness) function determines the environment within which the solutions “live.” Here, optimal solutions emerge through the evolution of the population, which takes place after the repeated application of some operators mimicking well-known natural phenomena: selection for reproduction, recombination, mutation, and selection for replacement. In reproduction, parents for the next generation are selected with a bias towards higher fitness. Parents then reproduce, and offspring (new candidate solutions) is generated through recombination and mutation. Recombination acts on the two selected parents

(candidates) by swapping some of their building blocks (genes), resulting in one or two children. Mutation acts on one candidate alone and results in a new candidate. Finally, the new candidates compete with old ones for their place in the next generation (survival of the fittest).

3. Individual and social learning in crime

The social interaction and learning aspects that underlie criminal activities were investigated in Sutherland's seminal work [36] in which the differential association theory was proposed. This theory advocates that interaction with others who are delinquent increases the likelihood of someone becoming and remaining a delinquent. That is, peers can play a crucial role in the development of values and beliefs favorable to law violation. In this theory, Sutherland elaborates nine postulates, among which three are particularly relevant for our study:

- Criminal behavior is learnable. This means that behavior toward crime is not an inherited trait and is not something to be acquired only by wish²;
- Criminal behavior can be especially learned through the interactions one establishes with other persons, typically through a verbal communication process; and
- The main part of the learning of criminal behavior occurs within intimate personal groups.

Recently, different perspectives on the study of crime in human societies have appeared, capitalizing from the theoretical resources made available in the area of social network analysis [33] and information systems [12,39]. Some works have focused on characterizing the impact of social network topologies (*viz.* scale-free and small-world settings) on the development and growth of special types of criminal activities, such as those related to narcotics [26]. By other means, Calvó-Armengol and Zenou [10] have studied, through a game-theory stance, how the levels of criminal activity in a given territory are influenced by the competitive-cooperative relationships established by delinquents dwelling it. The authors concluded that the various equilibria produced by the game, representing different numbers of active criminals and their levels of involvement in criminal activities, are only driven by the geometry of the social links connecting the delinquents. Other researchers, such as [24], have verified the influence of social relations on the motivation of young boys to commit petty crimes. Basically, the general conclusions taken from these studies, which corroborate those of Sutherland [36], are that criminal behavior depends not only on individual incentives but also on the behavior of the individual's peers and neighbors. In other words, an individual is more likely to commit crimes if his/her peers usually commit crimes.

Finally, another important result coming from works investigating social network models within the context of criminology is that social networks come to be a natural way of explaining the concentration of crimes per area. Crime data analyzed from different regions, and even different countries, usually reflect the fact that there are huge spatial (and also temporal) variations in the crime rates among different cities and among different regions in a city. In this regard, Glaeser et al. [21] show that less than 30% of the spatial variation of crime (both inter- and intra-city) can be explained by differences in local attributes. The remaining 70% can be explained by social interactions, which means that the agents' decisions about crime are somewhat positively correlated. The authors also show that the impact of social relations is greater in thefts, burglaries, muggings, and robberies (*i.e.* crimes against property) than in homicides. It is important to note

² Research on different domains [8] advocates that even social traits are inherited as a genetic process. However, what is important for our purpose here is the fact that social behavior has a learning component.

that these research works are not mutually exclusive with others interested in the social aspects of crimes. Particularly, it is worth mentioning the convergence with the ideas coming from the Routine Activity Theory [13], which emphasizes the influence that opportunities have in determining the spatial patterning of crime. The social relations are not dissociated from the environment in which the agents are inserted, nor are they a consequence of the daily activities exerted by such agents.

4. Simulating crime and police patrol

One of our claims in this paper is that ant systems, augmented with social network concepts, comprise an adequate strategy for modeling criminal behavior. There are several reasons that ratify this choice. One is that criminals prefer to commit crime in locations known to be vulnerable, with high payoff, etc. In other words, their choice considers their preference and knowledge about the crime points. The link here to ACO is that, according to this approach, ants always choose their next location in the environment (the place they move toward) biased by a mechanism (the pheromone marks) that intuitively complies with the notion of preferential attachment.

Another interesting feature that ACO offers our purposes is that it includes concepts intrinsically related to the notion of 'collective.' We have already emphasized that characteristics perceived at the social level belong to groups of individuals and not to the individuals themselves. In ACO, ants perform their local search tasks without dictating the whole colony's behavior, which, in turn, is recognized as an emerging result coming from all these local activities.

In particular, our crime simulation model is composed of three basic agents: guardians (the police team), targets, and criminals. The model abstracts out ants as criminals that make decisions based on: (i) their private experience with each possible point of attack (*i.e.*, the personal level of attractiveness to these targets); (ii) the distance from their current place to each possible point of attack; and (iii) the experience shared by other criminals in their community with regard to each possible target. We assume that criminal agents take part in communities (that is, pertain to social networks) and share their knowledge about crime targets with their peers.

4.1. The police team

There is a set of police teams available, each one associated with a monitoring route passing through certain locations of the urban territory (*i.e.*, targets) being considered. A police patrol route of length n is defined as a set $R_t = \{P_{t1}, P_{t2}, \dots, P_{tm}\}$ each component of which is a triple $P_t = (T_g, \Delta_t, P)$, where T_g is the target, Δ_t is the interval of time the police team remains in the target, and P is the daily period (patrol shift) the routes refer to. Thus, P can assume one of the following values: morning (6 a.m. to noon), afternoon (noon to 6 p.m.), evening (6 p.m. to midnight), and night (midnight to 6 a.m.). There is no distinction, in terms of skills, between the police officers allotted to the different police teams. Different teams may be associated with different-length routes, which, in turn, can overlap and/or share common points of surveillance.

4.2. Targets

The locations to be patrolled are referred to as targets, which can be differentiated with respect to the type of commercial/entertainment establishment they represent (*viz.* drugstores, banks, gas stations, lottery houses, and malls). Targets are distributed in a geographic area, which, in turn, is represented by a digitalized map of the region. Mobile targets (such as people, cars) are not modeled, since we have concentrated our study on crimes against property. Targets have a state of vulnerability that can be either active or inactive. A vulnerable target means that it is perceivable by a criminal.

Otherwise, it would not take part in the set of choices of that criminal. Each target has a probability of being vulnerable, which follows the temporal distribution of real crime data for the associated target type. In doing so, we are modeling a control parameter that allows reproducing the pace of crimes per type as it happens in real life.

The temporal distribution of crime events varies on daytime basis. We have modeled this variation in four periods of six hours each, in the same way as for the police teams' patrol shifts mentioned before. For each period and type of target, a value for a configurable parameter, λ , must be determined at the beginning of the simulation in order to define the pace of occurrence of crimes. For instance, at evening, drugstore robberies may follow a distribution based on a given value for λ ; whereas, during daylight periods, the crime temporal distribution might shift, achieving values four times higher for λ . At any simulation tick, at least one target is made vulnerable in accordance with the temporal distribution associated with its related type.

4.3. Criminals

There is a set of criminals representing the agents that frequently try to commit crimes. Each criminal is endowed with a limited view of the environment, measured in terms of a radius in meters. Criminals have one or more points of departure that we call 'gateways.' Such points of departure represent places where criminals are likely to start out, e.g., their residences, metro stations, bus stops, etc., before committing crimes. It is also assumed that, at the end of each day, each criminal always returns to the initial gateway, something that does not depend upon the number of crimes he/she has committed on that day.

It is worth remembering that, in our model, the potential number of crimes (those that occur and those to be avoided) per type of target depends basically on the values of the parameter λ since they define, at each moment, the parcel of the targets that is made vulnerable. Therefore, as in each simulation tick only one crime event is possible to occur, the number of criminals does not seem to be very relevant for achieving a certain level of generated crime events. Yet, it is worth noting that the number of criminals is indeed an important control parameter to our model, since communication among criminals exists and directly influences their final decision about the next targets to choose.

Target selection is probabilistic (see next section for a formalization of the adopted strategy) based on the target vulnerability, distance, and the criminal's experience. The shortest period of motion, considering all criminals, is taken as a reference, so that the criminals are allowed to move only during this time period. Note that only vulnerable targets are considered in the target selection process. Finally, the decision whether or not to commit a crime is made based on the existence of one or more police teams within the radius of the criminal's sight [35]. If the offender decides not to commit a crime, then he/she will select a new target to approach, leaving the current location. Otherwise, we assume that a crime will be committed and another target will be selected subsequently.

4.4. Learning ability of the criminal

Criminal behavior has a learning component that exploits the offender's own experience with each target in conjunction with the information coming from other criminal agents. The success rate of individual agents is computed as the ratio of the number of successful crimes to the overall number of crimes attempted in their lifetime. Criminals form communities wherein hints are shared. The agents in a community communicate with each other at the end of a day period. Due to the interconnection of the communities, such hints could be relayed to other criminals in other communities, and the rate at which this happens depends directly on the topology of the network of communities.

In this study, we have considered a scale-free topology in which nodes with the highest number of acquaintances are elected hubs [4,26].

4.5. Swarm-based criminal agent behavior

In our model, each criminal has three possible actions: commit a crime, not commit a crime, and move to a certain location. In order to reach a decision whether or not to commit a crime, criminals make use of a probabilistic formula, given as Eq. (1), which is adapted from the context of ant-based swarm systems [14,15]. In this equation, p_{cn} represents the probability of a criminal agent, c , choosing a specific target of the environment, n :

$$p_{cn} = \frac{[\tau_{cn}]^\alpha \times [\phi_{cn}]^\beta}{\sum_{\forall p \in N} [\tau_{cp}]^\alpha \times [\phi_{cp}]^\beta}. \quad (1)$$

Here, τ_{cn} represents the learned experience of criminal c with relation to target n , whereas N is the set of all targets the criminal c considers while deciding where to commit the next crime. The other parameter, ϕ_{cn} , denotes a static value (not learned) that represents the inverse of the distance between the current location of criminal c and that of target n ; we assume that the criminal has the knowledge necessary to localize the closest exemplar target on the map. Empirical evidence [6,9] suggests that many criminals do not travel great distances in order to commit a crime. Most crimes against property are committed in neighborhoods near, but not too close to, the criminals' residences. Then, we decided to model the probability of a criminal's going toward a target as the inverse of the distance between that target and the criminal's current location (a particular gateway).³ The parameters α and β in Eq. (1) are employed to balance the importance of τ_{cn} against ϕ_{cn} . As ϕ_{cn} is the inverse of the distance, when β increases, the learning factor becomes more important.

Eq. (1) is applied with respect to all currently-vulnerable targets, and the decision is made by the criminal c regarding which target to attempt the next crime at. Once the target is chosen, the criminal will attempt to commit a crime; this is captured in the discussion that follows by CT_{cn}^p , which stands for the number of crime attempts for a criminal c at a specific target n in a period p . The number of attempts can be subdivided into:

- successful attempts (crimes effectively committed), denoted by CO_{cn}^p again representing a local (at a given target) counter, for a period p ; and
- attempts that were prevented by the police.

Regarding the learned experience factor at a moment t in the simulation, τ_{cn}^t , this can be calculated as

$$\tau_{cn}^t = \mu \times \phi_{cn}^t + (1 - \mu) \times [S_{cn}^t], \quad (2)$$

where

$$\phi_{cn}^t = \rho \times \phi_{cn}^{t-1} + (1 - \rho) \times \phi_{cn}^{ld}, \quad (3)$$

$$S_{cn}^t = \sum_{\forall k \in S(c)} (\tau_{kn}^t). \quad (4)$$

Eq. (2) indicates that the learned experience factor at a time t , τ_{cn}^t , is given as a function of two terms, ϕ_{cn}^t and S_{cn}^t representing, respectively, the private and collective experiences of c with respect to target n . In Eq. (3), the private experience is based on the division of number of occurred crimes by the total of attempted crimes for the entire simulation period until then (represented by the period $t - 1$), ($\phi_{cn}^{t-1} = CO_{cn}^{t-1} / CT_{cn}^{t-1}$) and the same division considering the agent

³ We do not consider that the criminal's home is a gateway.

experience taken from the last day (ld), ($\phi_{cn}^{ld} = CO_{cn}^{ld}/CT_{cn}^{ld}$). From this, a negative feedback factor ρ is applied, which represents the level of forgetfulness of a criminal's past experience with regard to target n ; that is, the extent to which the agent considers the experience taken from the last day (period ld) in relation to the experience it already has with regard to target n , for the entire simulation period until then ($t-1$). It is worth noting here that, in terms of ant-based swarm systems [15], the parameter ρ is used to represent a pheromone evaporation rate. Conversely, in our criminal model, this parameter represents the rate at which agents forget past crime events. Hence, given that τ_{cn}^t represents the level of confidence of a criminal with a given target, we emphasize that at every day (or every fixed interval of time), the agents forget a little about their previous experience and are more influenced by the recent one.

Eq. (2) also shows that an initial succession of failures will lose its influence over the agent's lifetime.

Eq. (4) captures the experience shared by the agent's peers with respect to a given target. The importance of this social learning factor is controlled by the parameter μ in Eq. (2). Note that the social network related to the criminal agent c is indicated as $S(c)$ in Eq. (4). In addition, every day the agent's own experience is weighted against the experience of all the others pertaining to the agent's social network. It is also important to discuss the information conveyed in Eq. (4). The value of ζ_{cn}^t represents the consensual judgment the acquaintances of c have with respect to the target being considered (n). In other words, all contacts of c tell it separately what they think about target n . Such hints are then aggregated.

5. Configuring gateways with genetic algorithms

As mentioned before, in our simulation model, a control parameter that needs special attention is the place from where each criminal starts out to commit crimes every day. Examples of these initial locations—here called gateways—are bus stops, subway stations, and slums. As there is usually no real data or theoretical model to help one configure those gateways in a crime simulation model, we have decided to cast this task as a combinatorial optimization problem, i.e., the problem of assigning criminals to gateways.

More formally, let $G = \{G_i, i = 1, \dots, N_g\}$ be the set of gateways and $C = \{C_j, j = 1, \dots, N_c\}$ be the set of criminals under consideration. The goal is to allocate each C_j to a G_i in a way that a quality measure F , somehow related to the aim of the simulation model, is maximized.⁴ In this allocation process, any gateway can be assigned to a criminal and all criminals must be allotted to one, and only one, gateway. Besides, more than one C_j can be appointed to a given G_i (that is, we have not imposed any limit over the number of criminals assigned to a gateway).

Since this assignment problem is combinatorial in nature, the number of feasible gateway configurations is an exponential function of the number of possible gateways. Therefore, in order to cope with this task, we have resorted to a modified genetic algorithm model, named *Clearing* [30], which makes use of the notions of species and niches [17]. According to this approach, the population of individuals (i.e., solutions to the problem in sight) is adaptively segregated into species, each one exploiting a niche (particular region) of the search space. The idea is to segregate and preserve different species in order to exploit different high-quality niches, i.e., regions of the search space associated with (quasi-)optimal solutions.

In our customized GA instance, each chromosome represents a possible gateway configuration, that is to say, a valid assignment of criminals to gateways. Each one of the N_c genes of a chromosome receives a value i ($i = 1, \dots, N_g$) representing one of the possible

gateways. For evaluating a given individual, the GA interacts with the simulation model, passing to it the gateway configuration represented by that individual as input and receiving from it the corresponding fitness value. As the fitness function, we have adopted Pearson's chi-square (χ^2) test [22], whereby it could be possible to compare the spatial distribution of crime events produced by the simulation with that underlying real data. More precisely, at the end of each simulation run, a comparison between real data and simulation data is conducted in terms of the number of crimes effectively committed and the places where these crimes were committed. Therefore, the result of the χ^2 test provides a quantitative measure for indicating how properly the simulation model—adjusted with a particular gateway configuration—can reproduce the real crime distribution pattern. More formally, the χ^2 test is

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Where O_i = the frequency of real crime for an area i ;

E_i = the frequency of simulated crime (the expected) for the same area i ;

n = the number of areas.

It is important to point out that, due to the stochastic nature of the simulation model, each chromosome is evaluated N_s times ($N_s > 1$). The fitness value of a chromosome is thus obtained by averaging the χ^2 test values achieved in all N_s simulation runs. The definition of the niches is dynamic and is given in accordance with a certain similarity measure, which in our case is calculated over the decoded solutions and takes into account aspects of the domain under investigation. Preliminary experiments have evidenced that gateways close to each other on the map display quite the same level of influence over the targets around. Therefore, the similarity measure we have adopted to compare two GA individuals takes into account the physical distances of their constituent gateways. Thus, two individuals will likely pertain to the same niche if the overall sum of the paired distances between the gateways that belong to them is below a certain threshold, σ_{clearing} , which in the context of Clearing is known as the clearing radius [30].

Besides σ_{clearing} , another control parameter that needs special calibration is k , the maximum allowed number of individuals per niche. Both parameters delimit how many niches can be maintained as well as the granularity with which different niches can be discriminated. The final outcome of the algorithm is, thus, a list with the best individuals of each niche produced in the last generation, which—in our case—are expected to represent different first-class criminal-gateway assignments.

Other components of our GA instance come as follows. The variability of the population is provided by making use of one-point crossover and simple mutation, whereas the roulette wheel method is used as selection operator for both reproduction and population replacement [17].

6. Experimental methodology and results

In order to evaluate the suitability of our bio-inspired crime simulation model, experiments have been carried out over an artificial urban environment (referred to here as the simulation map) that mimics a well-known neighborhood of Fortaleza, a 2.5-million inhabitant metropolis in the northeast of Brazil. As shown in Fig. 1, we have taken into account all existing 172 fixed targets in that neighborhood, which include drugstores, gas stations, lottery houses, banks, and shopping malls. The location of these targets on the map represents with fidelity their actual geographical distribution across the real urban space.

⁴ In a typical assignment problem involving agents and tasks [29], each assignment of a task to an agent has an associated cost. In our case, we assume that the assignment cost is the same.

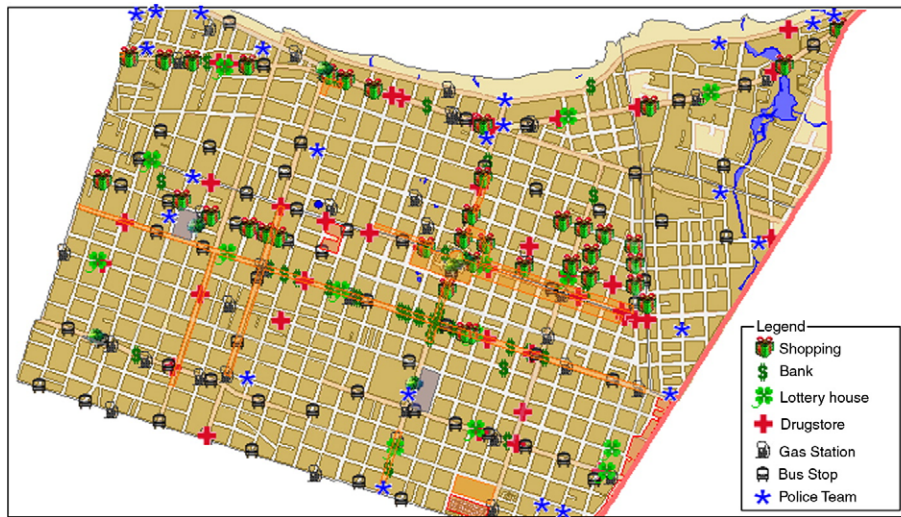


Fig. 1. Simulated environment representing a well-known neighborhood in the city of Fortaleza, Brazil.

In all, twenty police teams were made available to patrol the area. Fig. 2 depicts the number of routes (each route corresponding to one police team) allotted per day during the whole simulation period. As we can observe, the number of police teams effectively patrolling the area varied throughout the period, following a schedule that is similar to that typically used by the police department for that region. So, the maximum number of teams allowed to patrol was twenty (at the beginning of the simulation) whereas the minimum was four (at day 47). When more than one area had to be patrolled in the same patrol shift, the time of stay at each target was divided equally. The places to which the police teams were assigned can be seen in Fig. 1 (represented by a blue *).

Other control parameters related to the multiagent simulator come as follows. The number of criminals (N_c) adopted in our experiments was set to sixteen, as we have empirically verified that with this fixed contingent it would be possible to reproduce, with reasonable fidelity, the number and frequency of crimes observed in real life. In all simulation runs related to the experiments reported here, we have adopted a scale-free topology for the social network of the criminals, since our previous experiments [18] have also indicated that this topology is adequate to reproduce a Zipfian spatial distribution of crime events.

Moreover, it is pertinent to mention again that the temporal distribution of crime events in our simulation model follows an exponential distribution profile, as identified by [11] from real crime data analysis. Thus, the probability of a target being vulnerable follows an exponential distribution and consequently drives the crime

occurrences at that temporal pace. It is also worth remembering that the crime exponential distribution depends also on the period of the day at which the crime occurs. The values of λ for each daily period were obtained after careful analysis over almost four months of collected real crime data and are computed as the inverse of the average time measuring the frequency of occurrence crimes in hours, shown in Table 1.

After tuning the multiagent simulator, we conducted experiments with our GA approach for dealing with the criminal-gateway allocation problem. For such a purpose, the population size adopted was 50 individuals, the same number of the generations used for evolving the population. On the other hand, the crossover and mutation rates employed were of 95% and 5%, respectively, whereas the parameters related to the Clearing model were $\sigma_{clearing} = 7000$ and $k = 10$ (set after some manual fine-tuning). The number of simulation runs for each individual, N_s , was set as three. The results indicating how close the crime events generated by the simulation were to those observed in real life are presented next, following two sorts of analyses: statistical and visual. Regarding the first type of analysis, Fig. 3 presents a bar chart depicting the chi-square distance values between 10 crime distributions achieved through simulation and the real crime distribution observed in the neighborhood considered. These distance values are computed by means of a pairwise comparison taking into account all targets.

The red line on the chart indicates the limit below which the difference (distance) between the two crime distributions (real and simulated) is not statistically significant. This limit value was set after

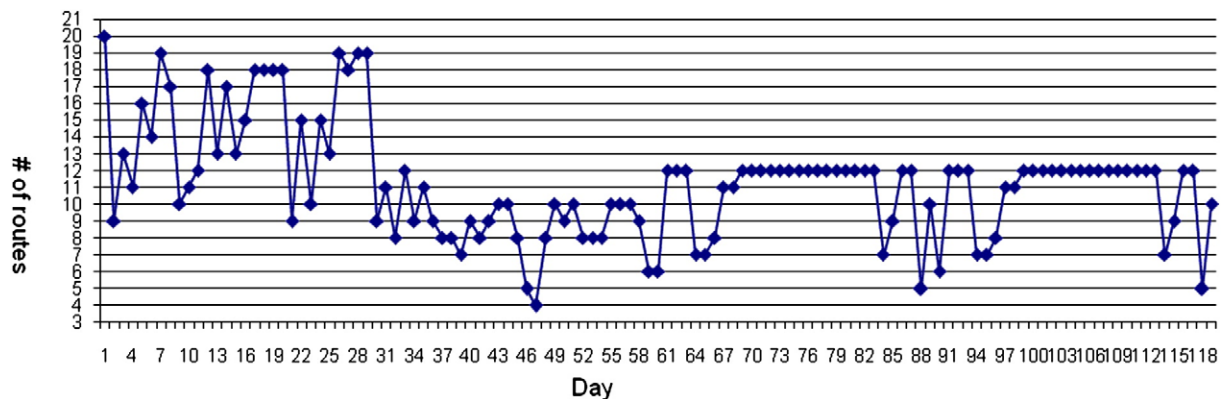


Fig. 2. Distribution of patrol routes during the simulations.

Table 1
Average frequency, in minutes, between crimes involving each target per period of the day.

	Bank	Lottery house	Gas station	Drugstore	Mall
Night	–	–	0.73	0.26	0.21
Morning	7.25	0.56	2.51	0.40	0.08
Afternoon	–	1.34	2.57	1.36	0.08
Evening	–	1.75	0.93	0.59	0.09

resorting to a χ^2 distribution table [22]. Each one of the ten chi-square values was achieved by running the simulation model with a gateway configuration produced by an individual discovered by the GA engine. As ten niches were discovered, ten best niche representatives were selected to configure the simulation in terms of criminal–gateway allocation. As the chart evidences that all the distance values achieved remain below the red line, one can conclude that the distribution of crimes generated by the particular gateway configuration represented by each individual is statistically equivalent to the real distribution. Fig. 4 shows kernel maps that allow one to visually inspect how close the two types of crime distribution (real and simulated) really were.

It is possible to observe that the formation of clusters of crime events is very similar between the map in (A) describing real data distribution, and those in (B), (C) and (D). Worth noting is the fact that the hot spots identified by simulation were indeed very close to those happening in real life; conversely, in most of the other zones of the map, the crime events occur in a lesser magnitude, characterizing the actual formation of a power-law distribution.

The conclusion reached from a visual analysis of Fig. 4 can also be confirmed by the results obtained with geospatial statistics methods [28] such as K-function (Ripley) [32], which permits an assessment based on the analysis of Euclidean distance between the points of the set, if there is an indication of concentration of crime occurrences. For a more suitable comparison of the two distributions (real and simulated), we considered a variant of the K-function that can be applied to non-stationary processes [3]. In this case, the lambda (λ) considered for the theoretical distribution of reference is the intensity expected for the occurrence of real crimes. Fig. 5 shows a graph containing the $K(r)$ values, estimated according to the patterns of points on the theoretical curve (in red) and the observed curve (the simulated data for the best individual of the population), here already calculated with the border adjusted in order to reduce the likely bias

of the estimate. We can observe that the simulated data follows the same pattern of the theoretical one indicating a similar level of clustering.

Spatial correlation techniques help to measure the proximity in (two-dimensional) space between observations of the same phenomenon. In other words, they indicate whether the events are correlated. Local spatial autocorrelation statistics provide estimates disaggregated to the level of the spatial analysis units, allowing an assessment of the dependency relationships across space. We have applied Global and Local Moran's I tests [2] to identify the assessment of significant local spatial clustering around an individual location in real and simulated data. Global Moran's indicate a weak positive correlation of crimes in real (0.1438) and simulated data (0.1250). Fig. 6 shows the choropleth map (called the significance map) showing the locations with a significant Local Moran statistic as different shades of green, depending on the significance level for simulated and real data. The simulation data were obtained from a simulation in which the configuration of gateways was defined from the best individual found by the GA. The maps have demarcated a typical spatial unit used in Brazil: the 'censor region' (an area that corresponds to around 300 families). It is possible to see that in general the same regions were identified as significant according to Local Moran statistic ($p=0.05$).

Finally, we validated the simulation data by verifying if it follows Zipf's Law. A Zipfian distribution is most easily observed by scatter plotting the data in a log–log plot (rank order against frequency). If the points were close to a single straight line downwards, this means that the distribution follows Zipf's Law. Fig. 7 shows the crime distribution per target produced by a simulation run in which criminals start out in the gateways produced by the final best individual of the whole population. The targets are ranked in descending order, meaning that the first target has the highest number of crimes. We have also computed the linear regression coefficient R^2 for each distribution to numerically measure the distance between the simulated data and a linear curve. The coefficient of determination (R^2) measuring the distance between the simulated data and a linear curve, is 0.85—a mark very close to that calculated for real crime data (0.89). In [18] we make a detailed analysis that indicates that the Zipf factor is somewhat correlated with the social factor. The three simulation variants with social communication show high regression values (≈ 0.85) indicating that a Zipfian profile is indeed present (we consider those values as Zipfian-

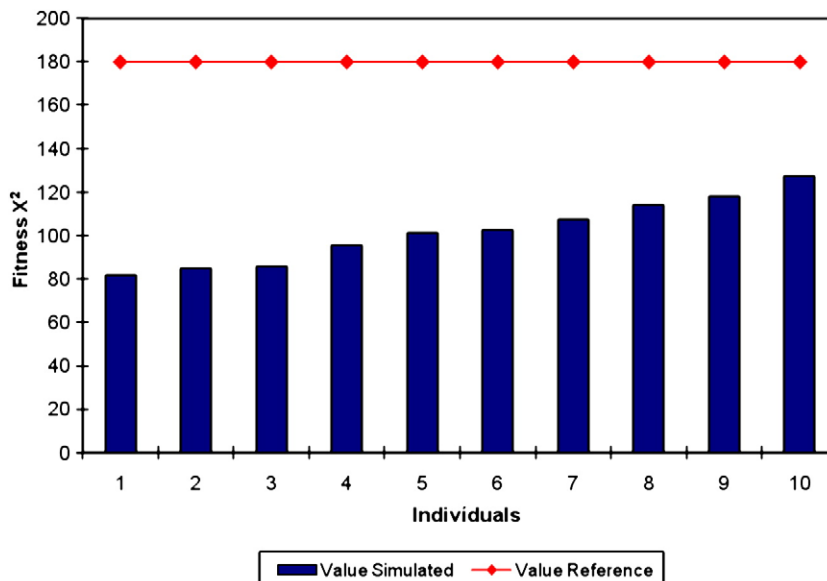


Fig. 3. Distance values between simulated and real crime distributions—each simulation was configured with an individual (criminals–gateway allocation) taking from a different GA niche.

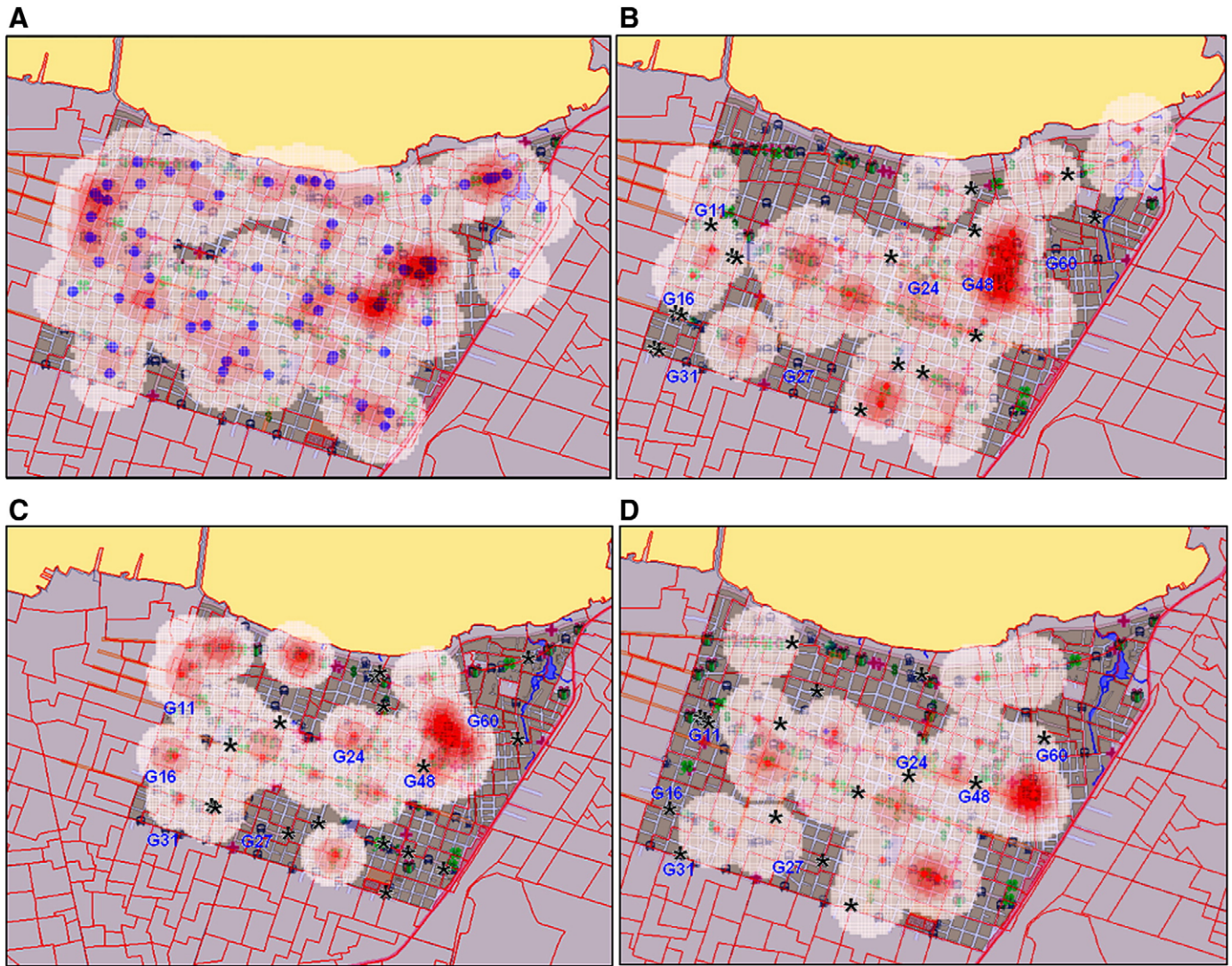


Fig. 4. Maps showing real crime data (A) and crime events generated in the simulation with gateway configurations produced by the best three individuals of the population (B, C, and D).

like because it is a stochastic phenomenon and uses the average of simulations. Moreover, the slope values were close to 1, as in a typical Zipfian distribution). The regression factor is sensibly reduced when-

ever there is no criminal communication. By varying the topology of the social network, we could analyze how sensitive the model is to a particular topology.

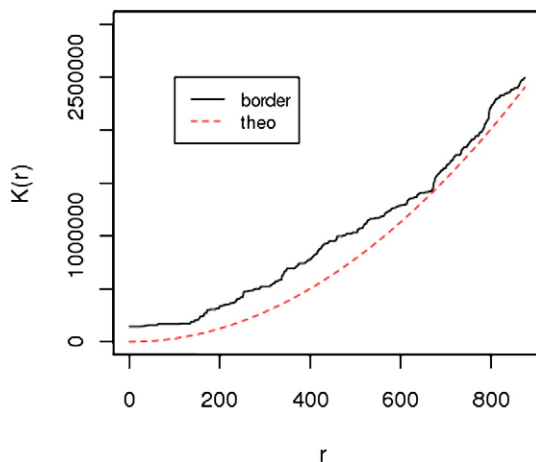


Fig. 5. Global spatial analysis using Ripley's K-function.

7. Discussion

One of the particular advantages of using a GA to tune the proposed crime simulation model has to do with the possibility of applying it as a decision support tool. We understand that the results produced by the GA may help police experts in identifying potential gateways which deserve special attention in the future preventive surveillance. In this regard, we have conducted a qualitative evaluation of the results produced by our GA engine in order to reveal possible patterns existing among the final different solutions produced by it, representing alternative gateway configurations. More precisely, our idea here is to analyze how plausible the gateways found by the GA are.

In Fig. 8, we display a histogram reporting how many times each potential gateway was indeed selected by the GA engine to compose the ten best niche individuals produced at the end of its execution. Analyzing that histogram, it is possible to identify the gateways that were most frequently selected. Among them lie gateways G24, G48, and G60, which are the closest to the hot spots. Such result was already expected by us, since the distance to the targets is an

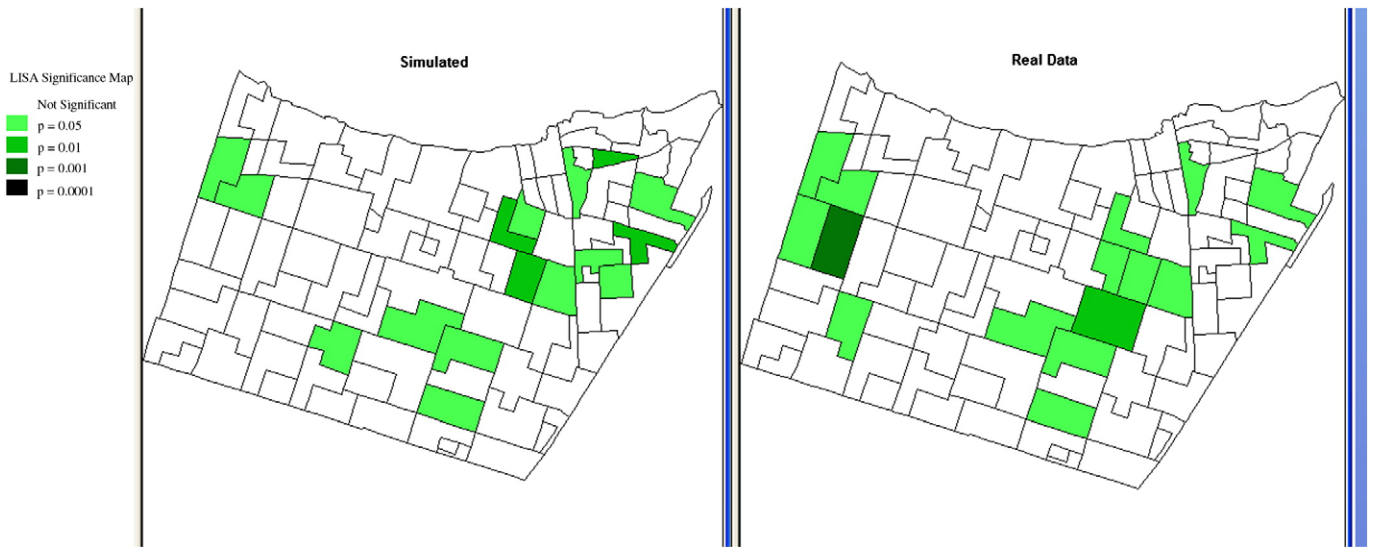


Fig. 6. Significance maps for simulated and real data.

important issue underlying the criminal behavior in our model. Our perception is that the possibility of discovering such frequent and plausible gateways (in the sense that they are close to the hot spots) is of fundamental relevance to law authorities, particularly while deciding their preventive patrol strategies in the regions where those gateways are located. Gateways close to hot spots are, of course, strong candidates to be chosen by the GA as being the most plausible

ones. It is important to point out that even knowing that the problem of finding a good allocation of criminals in gateways is far from trivial and cannot be made by hand following simple heuristics (such as to place criminals close to hot spots). In a complex configuration, there are several hot spots and there are several gateways close to these hot spots. Moreover, one does not have any information about how many criminals should be allotted to each one of them.

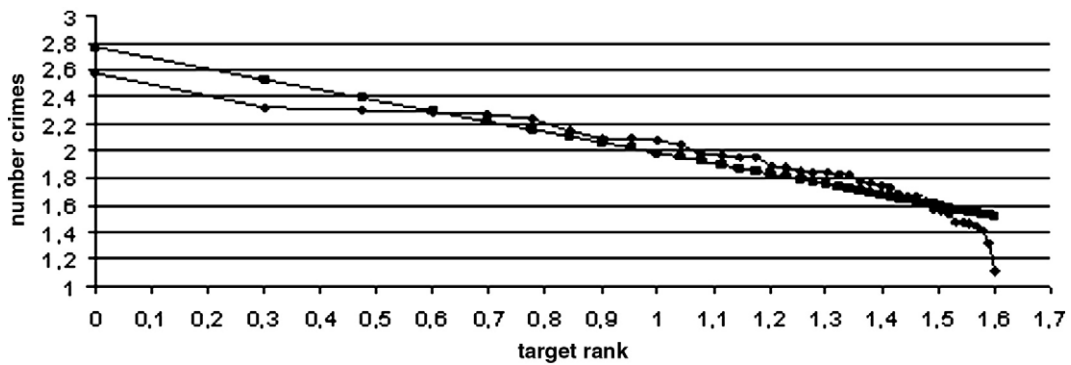


Fig. 7. Crimes per target following Zipf's Law distribution.

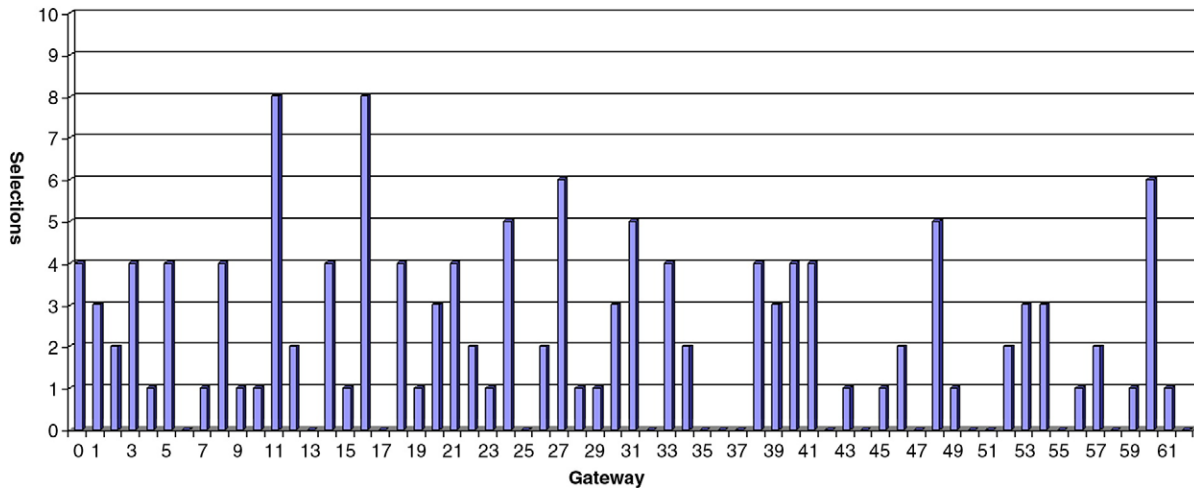


Fig. 8. Selections per gateway.

One of the limitations we have already spotted in our approach is that it is based on the use of historical real crime data to define the pace of crimes (parameter λ). This implies that, during a whole simulation run, the potential number of generated crime events will always be the same. In other words, the current version of our simulation model is not capable of identifying changes in the pace of crime occurrences if that change actually occurs in real life. This may turn out to be a big issue to be dealt with inasmuch as the simulation time increases. However, as we plan to use the model for discovering effective patrol routes on a frequent basis (possibly every week), the simulation time necessary shall not be very long.

Another limitation (closely related to the previous one) refers to the fact that, currently, the number of criminals used in the reported experiment is constant along a given simulation run. That is, no mechanisms (such as the simulation of arrests, deaths, etc.) exist to implement variability in the number of criminals. The rationale behind our choice is again related to our ultimate goal of finding good police patrol routes. Considering constant the number of criminals means that crime reduction is only attainable by preventing a potential criminal from acting. In a certain way, such decision is positive, since it allows our model to be assessed in the worst of the scenarios.

A final note worth pointing out is that most of the research conducted so far on crime simulation has not made systematic use of optimization tools (either heuristic or not) to fine-tune their proposed models [27,38]. Our perception is that this mostly occurs because the level of refinement underlying these models is not as detailed as the one proposed here. We understand that the calibration of models that try to mimic reality is of paramount importance if one wishes to obtain significant results, and that the GA-based assignment approach as employed is a step in such direction.

8. Conclusion

The simulation of criminal activities in urban environments is an asset for decision-makers seeking to find preventive measures. Law-enforcement authorities need to understand the behavior of criminals and their response to public-safety measures and policies. In this paper, we have shown that ant-inspired systems [7,15] constitute an adequate metaphor for modeling criminal behavior. By exploiting this approach, we could properly model the effect criminals' preferences have when they are committing their crimes and also the way these preferences depend on past experience and social communication. In other words, the ant-based approach allows for crime target selection via a mechanism of preferential attachment [25]. As a consequence, the proposed simulation model can yield a spatial distribution of crime events that agrees with real data.

We understand that the results presented in this paper should serve as a preliminary foundation for future investigations intended to cope with some particular issues related to the following questions: Does the non-reestablishment of the Power Law translate into a decrease in crime? We know crime follows a Zipfian distribution, but is this a necessary and sufficient characteristic for the existence of crime as a phenomenon? Can Zipf's Law persist through time even after a series of preventive strategies takes place? If not, how long does it take for one to notice the reestablishment of this law? Is it somehow possible to avoid such reestablishment?

By other means, in parallel to the present work, we have been investigating a hybrid methodology that combines genetic algorithms and multiagent systems to assist police officers in the design of effective police patrol routes. Our idea is to uncover strategies for police patrolling (more precisely, police patrol routes) that could cope well with the dynamics of crime when criminal agents are capable of learning "on the fly." In this context, we plan to provide satisfactory answers to the following questions: How far from the optimal patrolling routing strategies are those actually adopted by human police managers? How complex do such optimal patrolling routes need to be in terms of their

total lengths and urban area coverage? Preliminary results [31] have shown that such a hybrid approach is also promising.

References

- [1] R. Akers, M. Krohn, L. Lanza-Kaduce, M. Radosevich, Social learning and deviant behavior: a specific test of a general theory, *American Sociological Review* 44 (4) (1979) 636–655.
- [2] L. Anselin, Local indicators of spatial association – LISA, *Geographical Analysis* 27 (1995) 93–115.
- [3] A. Baddeley, J. Møller, R. Waagepetersen, Non and semiparametric estimation of interaction in homogeneous point patterns, *Statistica Neerlandica* 54 (2000) 329–350.
- [4] A. Barabási, *Linked: The New Science of Networks*, Perseus Publishing, Cambridge, 2002.
- [5] A. Barabási, R. Albert, Emergence of scaling in random networks, *Science* 286 (1999) 509–512.
- [6] R. Blocka, A. Galary, D. Brice, The journey to crime: victims and offenders converge in violence index offences in Chicago, *Security Journal* 20 (2007) 123–137.
- [7] E. Bonabeau, M. Dorigo, G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, Santa Fe Institute Studies in the Sciences of Complexity Series, Oxford Press, 1999.
- [8] R. Boyd, P.J. Richerson, *The Origin and Evolution of Cultures*, Oxford Press, 2005.
- [9] P. Brantingham, P. Brantingham, Environment, routine, and situation: toward a pattern theory of crime, in: R. Clark, M. Felson (Eds.), *Routine Activity and Rational Choice*, Transaction Books, vol. 5, 1979, pp. 259–294.
- [10] A. Calvó-Armengol, Y. Zenou, Social networks and crime decisions: the role of social structure in facilitating delinquent behavior, *International Economic Review* 45 (3) (2004) 939–958.
- [11] T.M.L. Canção, *Alocação e despacho de recursos para combate à criminalidade*, Master dissertation, UFMG, Belo Horizonte (in Portuguese), 2005.
- [12] H. Chen, *Intelligence and Security Informatics: information systems perspective*, *Decision Support Systems* 41 (3) (2006) 555–559.
- [13] L. Cohen, M. Felson, Social change and crime rate trends: a routine approach, *American Sociological Review* 44 (1975) 588–608.
- [14] M. Dorigo, G. Di Caro, L. Gambardella, Ant algorithms for discrete optimization, *Artificial Life* 5 (2) (1999) 137–172.
- [15] M. Dorigo, T. Stützle, *Ant Colony Optimization*, The MIT Press, Cambridge, 2004.
- [16] A. Drogoul, J. Ferber, Multi-agent simulation as a tool for modeling societies: application to social differentiation in ant colonies, *Proceedings of 4th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, *Artificial Social Systems (MAAMAW)*, Lecture Notes in Computer Science (LNCS), vol. 830, 1992, pp. 3–23.
- [17] A.E. Eiben, J.E. Smith, *Introduction to Evolutionary Computing*, Springer, Berlin, 2003.
- [18] V. Furtado, A. Melo, A.L.V. Coelho, R. Menezes, M. Belchior, Simulating crime against properties using swarm intelligence and social networks, in: L. Liu, J. Eck (Eds.), *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*, Chapter XV, Hershey, Information Science Reference, 2008, pp. 300–318.
- [19] V. Furtado, E. Vasconcelos, A multiagent simulator for teaching police allocation, *AI Magazine* 27 (3) (Fall 2006) 63–74.
- [20] N. Gilbert, R. Conte, *Artificial Societies: The Computer Simulation of Social Life*, UCL Press, London, 1995.
- [21] E. Glaeser, B. Sacerdote, J. Scheinkman, Crime and social interactions, *The Quarterly Journal of Economics* 111 (2) (1996) 507–548.
- [22] P.E. Greenwood, M.S. Nikulin, *A Guide to Chi-squared Testing*, Wiley, New York, 1996.
- [23] M. Guntsch, M. Middendorf, Applying population-based ACO to dynamic optimization problems, *Proceedings of ANTS 2002*, Springer Verlag, 2002, pp. 111–122.
- [24] B. Houtzager, C. Baerveldt, Just like normal: a social network study of the relation between petty crime and the intimacy of adolescent friendships, *Social Behavior and Personality* 27 (2) (1999) 117–192.
- [25] H. Jeong, Z. Néda, L. Barabási, Measuring preferential attachment in evolving networks, *Europhysics Letter* 61 (4) (2003) 567–572.
- [26] S. Kaza, J. Xu, B. Marshall, H. Chen, Topological analysis of criminal activity networks in multiple jurisdictions, *Proceedings of the 2005 National Conference on Digital Government Research*, ACM International Conference Proceeding Series, vol. 89, 2005, pp. 251–252.
- [27] L. Liu, X. Wang, J. Eck, J. Liang, Simulating crime events and crime patterns in a RA/CA model, in: F. Wang (Ed.), *GIS and Crime analysis*, Idea Group, 2005, pp. 197–213.
- [28] W.L. Meeks, S. Dasgupta, Geospatial information utility: an estimation of the relevance of geospatial information to users, *Decision Support Systems* 38 (1) (October 2004) 47–63.
- [29] D.W. Pentico, Assignment problems: a golden anniversary survey, *European Journal of Operational Research* 176 (2) (2007) 774–793.
- [30] A. Pétrowski, A clearing procedure as a niching method for genetic algorithms, *Proceedings of IEEE Congress on Evolutionary Computation*, 1996, pp. 798–803.
- [31] D. Reis, A. Melo, A.L.V. Coelho, V. Furtado, GAPatrol: an evolutionary multiagent approach for the automatic definition of hotspots and patrol routes, in: J.S. Sichman, H. Coelho, S. Oliveira (Eds.), *Proc. of IBERAMIA/SBIA 2006*, Lecture Notes in Artificial Intelligence (LNAI), vol. 4140, 2006, pp. 118–127.
- [32] B.D. Ripley, *Statistical Inference for Spatial Processes*, Cambridge University Press. (1988).

- [33] J.P. Scott, *Social Network Analysis: A Handbook*, Sage Publications, 2000.
- [34] L. Sherman, P. Gartin, M. Buerger, Hot spots of predatory crime: routine activities and the criminology of place, *Criminology* 27 (1989) 27–56.
- [35] L. Sherman, D. Weisburd, General deterrent effects of police patrol in crime 'hot spots': a randomized, controlled trial, *Justice Quarterly* 12 (4) (1995) 625–648.
- [36] E. Sutherland, *Principles of Criminology* 4th ed, J. B. Lippincott Company, Philadelphia, 1974.
- [37] G. Weiss, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, The MIT Press, Cambridge, 1999.
- [38] Y. Xue, D. Brown, Spatial analysis with preference specification of latent decision makers for criminal event prediction, *Decision Support Systems* 41 (3) (March 2006) 560–573 Special issue: Intelligence and security informatics.
- [39] J.J. Xu, H. Chen, Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks, *Decision Support Systems* 38 (3) (2004) 473–487.
- [40] G.K. Zipf, *Human Behaviour and the Principle of Least-Effort*, Addison-Wesley, Cambridge, 1949.

Vasco Furtado is professor of computer science at University of Fortaleza (UNIFOR), Brazil, where he also leads a team of researchers in the Knowledge Engineering group that studies agent-based simulation and agent's explanation on the web. He is consultant of the State of Ceará IT Company (ETICE) where he has coordinated and developed R&D projects on the law-enforcement domain. Furtado holds a Ph.D. in Informatique from the University of Aix-Marseille III, France. He had a sabbatical in the Knowledge Systems Laboratory at Stanford University in 2006–07. Further information about publications and projects is available at <http://www.wikinova.com.br/vasco>.

Adriano Melo received his master degree at the University of Fortaleza (UNIFOR/MIA). His research interests include Multi-Agents Systems, Genetic Algorithms and Swarm Intelligence. He majored Computer Science at the University of Fortaleza in 2003 and received a diploma of Specialist in Information Technology from the Federal University of Ceará in 2005.

André L. V. Coelho was born on June 11, 1974, in Fortaleza, Ceará, Brazil. He received the B.Sc. degree in Computer Engineering in 1996, and earned the MSc and PhD degrees in Electrical Engineering (specialization in Computer Engineering) in 1998 and 2004, respectively, all from the State University of Campinas (Unicamp), Brazil. Currently, he is an adjunct professor affiliated with the Graduate Program in Applied Informatics at the University of Fortaleza (Unifor), Brazil, conducting research in computational intelligence, natural computing, metaheuristics, multiagent systems, machine learning, and data mining. He has a record of published papers on these themes. He is a member of ACM and IEEE.

Ronaldo Menezes is associate professor in computer science at Florida Institute of Technology. He received his BSc in computer science in 1992 from the University of Fortaleza, Brazil. He was awarded MSc (by research) in Computer Science in the field of Parallel Computing Models in 1995 from the State University of Campinas, Brazil. In 2000 he was awarded his Ph.D. in computer science at the University of York, UK. He is a member of many organizations such as ACM, SBC and AWC, and a senior member of the IEEE. He is currently serving as an elected official of the ACM SIGAPP (Special Interest Group in Applied Computing). He has published more than 40 papers in international events and journals. He is a program committee member of many conferences and workshops in Swarm Intelligence, Coordination and Complex Networks fields. His research interests are in the field of Swarm Intelligence and Coordination in Multi-agent Systems.

Ricardo Perrone obtained a MSc in Mechatronics at the Federal University of Bahia, Brazil, where is participating on team of researchers of the Distributed Systems Laboratory. He is a specialist in distributed systems and currently is associated researcher in the Knowledge Engineering group that studies agent-based simulation at University of Fortaleza (UNIFOR), Brazil. His areas of interest are: Applied Statistics, Real-time, Robotic, Artificial Intelligence and Distributed Systems.