Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE DÉFENSE

Production of preliminary traffic data for path planning in urban environment

From probability distributions to iterative shortest path algorithms

A. B. Guyard
L. Pigeon
DRDC Valcartier

R. Proulx
Thales Systems Canada

Defence R&D Canada – Valcartier
Technical Memorandum
DRDC Valcartier TM 2006-792
March 2008

Canada

Production of preliminary traffic data for path planning in urban environment

From probability distributions to iterative shortest path algorithms

A. B. Guyard
L. Pigeon
DRDC Valcartier

R. Proulx
Thales Systems Canada

Defence R & D Canada – Valcartier

Technical Memorandum

DRDC Valcartier TM 2006-792

March 2008

Author

A. Bergeron Guyard

Approved by

Stéphane Paradis
Section Head, Intelligence & Information Section

Approved for release by

Christian Carrier
Chief Scientist

Abstract

The SCIPIO Optipath module aims at providing path planning abilities in evolving urban environments. To effectively do so, the system has to take into account numerous factors (urban environment, sensor/humint force tracking, threats, etc.). Among those factors, traffic is an essential one. Traffic has a direct impact on how easy it is to maneuver in a city. However, traffic data can be hard to acquire. While it is possible to obtain maps of a city and to track enemy forces or threats, traffic data is very hard to collect. Attempts have been made using numerous probe cars over long periods of time (weeks or months). Such constraints cannot easily be dealt with in a deployment context. We therefore, must have a fast and efficient way to provide some usable preliminary traffic data. This paper presents a way to obtain such data. It is based on high-density attraction points (or hotspots), which dictate traffic behavior patterns.

Résumé

Le module SCIPIO Optipath a pour but de permettre la planification et l'exécution de déplacements dans un cadre urbain changeant. Pour ce faire, plusieurs facteurs doivent être tenus en compte (cadre urbain, capteurs/humint, positions, menaces, etc.). Parmi ces facteurs, la circulation présente sur le réseau routier en est un prépondérant. Il est possible de se procurer les cartes d'une ville ou de connaître le déplacement de troupes ou de menaces. Il est cependant moins facile d'obtenir des données de circulation. Des recherches ont été faites au moyen d'un grand nombre de véhicules sondés sur une longue période de temps (des semaines, voire des mois). Ces contraintes sont irréalistes dans une situation de déploiement. Il est donc impératif d'obtenir un moyen qui nous fournisse rapidement des données préliminaires de circulation. Ce document présente une solution à ce problème. L'approche proposée est basée sur l'étude de points d'attractions affectants le comportement de la circulation.

This page intentionally left blank.

Executive summary

The SCIPIO Optipath module aims at providing path planning abilities in evolving urban environments. To effectively do so, traffic data is needed. One may not take the same path at rush hour than in the middle of the night. It is important to know which road segments are busier at particular times. Over the years, the DRDC Valcartier SCIPIO research team has developed expertise working on routing problems in dynamic urban environments. Using this experience the team has put together a novel technique allowing for the production of preliminary traffic information.

The proposed solution allows us to extract relevant preliminary traffic data from human intelligence. Other techniques require a lot of time and resources (vehicles), both of which may not be available in a deployment context. This solution represents a fast and viable alternative that can be used within hours, making it desirable. The data we are able to extract is as accurate as the information it is derived from. To extrapolate the traffic information, we rely on the identification of different points of attraction, or hotspots, present in a city. The principle is that commuters will head towards an attraction point, for instance a big office building, at a certain moment of the day. At another time, drivers will leave the attraction point. This behaviour will impact traffic. Three general algorithmic and mathematical approaches modeling this phenomenon were developed, and are detailed in this document.

These new approaches provides useable preliminary traffic data. However, as useful and exact this preliminary data is, it is meant to be refined and validated through practical use. Work remains to be done in order to fuse the data we obtained from this technique and other information, which would be collected through actual use of the network. Deeper experimentation with this solution will allow the tailoring of the algorithms to better fit practical use.

A. B. Guyard, L. Pigeon, R. Proulx; 2008; Production of preliminary traffic data for path planning in urban environment; DRDC Valcartier TM 2006-792; Defence R & D Canada – Valcartier.

Sommaire

Le module SCIPIO Optipath offre la possibilité de planifier des déplacements dans un milieu urbain en constante évolution. Pour ce faire, il est nécessaire d'avoir une connaissance de la circulation présente sur le réseau routier. Il est évident que des routes différentes pourraient être empruntées à l'heure de pointe et dans les moments d'accalmie. Il importe donc de savoir quelles sont celles qui sont les plus achalandées à différents moments de la journée. Au cours des dernières années, l'équipe de recherche SCIPIO du Centre de RDDC Valcartier a développé une expertise dans la recherche de chemins dans des contextes urbains dynamiques, ce qui lui a permis de concevoir une nouvelle méthode pour obtenir des données de circulation préliminaires.

Cette méthode, basée sur le renseignement humain, permet d'extraire une connaissance initiale de la circulation présente sur le réseau routier. Les autres techniques requièrent du temps et des ressources (véhicules) qui ne sont pas nécessairement disponibles dans une situation de déploiement. Un moyen qui peut fournir des résultats utilisables en quelques heures est donc très souhaitable. Les données extraites seront aussi précises que le sera l'information dont elles émanent. Pour extrapoler l'information sur la circulation, on doit identifier des points d'attraction, ou hotspots, dans une ville. Cela s'appuie sur le principe selon lequel les automobilistes se dirigeront, à un certain moment de la journée, vers un point d'attraction, par exemple un gros édifice à bureaux, et, à un autre moment, ils quitteront ce point d'attraction. Ce comportement aura un impact direct sur la circulation. Trois approches algorithmiques et mathématiques permettant la modélisation de ce phénomène ont été développées et sont présentées dans ce document.

Ces nouvelles approches permettent d'obtenir des données de circulation préliminaires utilisables. Cependant, même si les données obtenues sont utiles et fidèles, elles doivent être raffinées et validées dans un contexte pratique. Il reste donc à fusionner les données obtenues au moyen de cette technique et l'information qui sera fournie en utilisant le réseau. De plus, l'expérimentation des différentes approches permettra d'ajuster les algorithmes pour mieux représenter le monde réel.

A. B. Guyard, L. Pigeon, R. Proulx; 2008; Prépopulation de bases de données de circulation en milieu urbain; DRDC Valcartier TM 2006-792; R & D pour la défense Canada – Valcartier.

Table of contents

List of figures

This page intentionally left blank.

1 Introduction

The SCIPIO Optipath module aims at providing path planning abilities in evolving urban environments [1][2]. To effectively do so, the system has to take into account numerous factors (urban environment, sensor/humint force tracking, threats, etc.). Among those factor, traffic is an essential one. In order for the Optipath module to be fully functional, it is essential to have some idea of the traffic present on the road network.

1.1 Extracting Traffic Data From Car Systems

Numerous studies have been conducted on this approach. A number of vehicles are used to detect the actual traffic situation. In other cases [3][4], vehicles are used as probes, and their speed is studied over the road network. From the data collected, segment traffic can be extracted. In some other cases [5], traffic data is collected through checkpoints. By assessing the time it took for each vehicle to drive over particular road segments, traffic data can be inferred.

All of these approaches are interesting. However, they share two common flaws. They require the use of many vehicles over long periods of time. These two constraints may be hard to satisfy in a deployment context. We are looking for a solution that would quickly provide useable preliminary data.

1.2 Theoretical Model

Before we go into a more detailed view of our solutions, it is important to set the theoretical context in which we are going to evolve.

In our research, the road network is represented by a directed weighted graph $(G = V,E)$. Each vertex represents an intersection, and each edge, a street segment. The weight of an edge is a function of time $w = f(e,t)$. Traffic or congestion level will affect the weight function of an edge e at a certain time t.

2 Probability Based Solutions

A traffic hotspot refers to a particular traffic attraction point in an urban environment. The following assumptions are made:

- Edges outgoing/incoming from/to a hotspot source/sink vertex are affected at the same proportional level if they lie at comparable distances from the source/sink

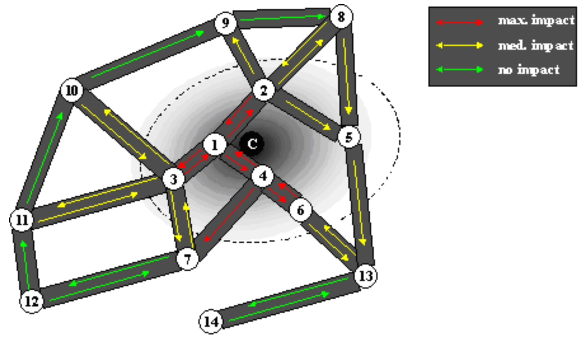

Figure 1: *Example of Hotspot on Road Network*

regardless of their respective number of lanes, i.e. their respective flow rate (and speed) is affected with the same proportion.

- The boundary of the hotspot influencial area is specified or drawn on the map by the user.
- A hotspot must have a reference point that lies within the hotspot area of influence.
- A hotspot reference point needs not be the geometric or mass center of the influence area, it is rather considered as the point where influence is maximum.
- A source or sink hotspot reference point (center of a particular traffic point of attraction) needs not be one of the network vertex, i.e. it can be any user specified location on the map.

2.1 Basic Solution

We will apply a distribution function that will affect the traffic congestion level within the hotspot, so that segments that are closer to the reference point of the hotspot are more affected than those that are farther away. We will discuss thereafter an incremental approach based on this idea which provides, at the lowest level, a quickly implemented solution and we then gradually consider enhancements that could refine it.

A quick first implementation would be based on the following premises:

- Segments affected will consist of all segments contained in the hotspot and all segments which intersect the hotspot boundaries, i.e. all segment arcs that have at least one origin/destination node lying within the hotspot area.

- Whether the hotspot is a source or sink will not change the way its segments are affected, i.e.
  – two-way segments will have both arcs affected in the same way
  – and one-way segments will be affected regardless of their direction

- The distances between the hotspot reference point and a segment edge will be computed as the smallest Euclidian distance between the reference point and the origin or destination vertex of the segment arc, provided that the corresponding vertex lies within the hotspot.

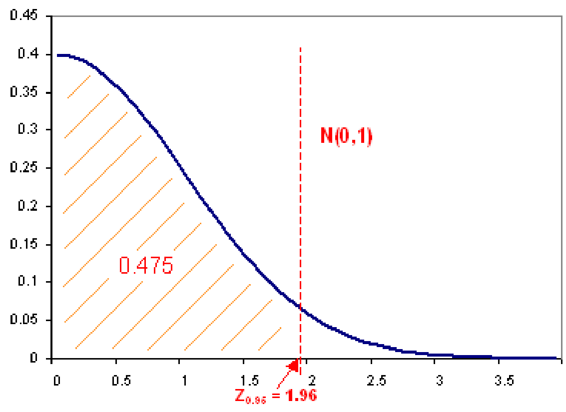

Figure 2: *Half-Normal Distribution*

The default distribution function used is the standard normal distribution $N(0,1)$. The user should be allowed to specify other parameters for the normal distribution. If other parameters $(\mu \neq 0, \sigma^2 \neq 1)$ are used, the distribution will be transformed into a standard normal distribution $N(0,1)$ for the sake of computations. Only the positive half-distribution of $N(0,1)$ need be considered (Figure 2). Since the distances are finite while the domain of the distribution is infinite, the distribution values will be truncated at 95% of the probability mass. Since the hotspot area need not be circular and have its reference point located at its center, the distance variable will be normalized, so that the impact at distance 0 is maximal (according to the half-normal distribution) and null on and outside the boundary of the hotspot. The actual variable is computed as the distance between the hotspot reference point and the edge divided by the length of the director radius passing through the edge reference vertex. Impact will be computed from the probability distribution function (PDF) of a standard half-normal distribution. Impact will be expressed as a real number $w \in [0,1]$ by normalizing it over the maximum probability distribution function (PDF) value, i.e. PDF(0) = 0.3989.

In the case where the reference point of the hotspot is located at a vertex (Figure 3), all segments issued from that vertex will have their distance computed as 0. It

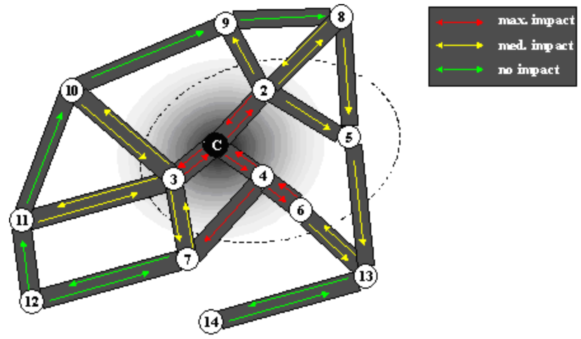

Figure 3: *Example of Hotspot with Reference Point on Vertex*

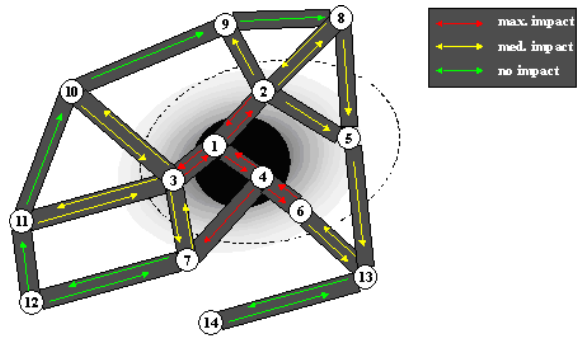

Figure 4: *Example of Hotspot with Area Reference Point*

is unrealistic to assume that the corresponding impact $(= 1)$ will imply that the speed on this segment is reduced to 0, thus a minimal threshold value will be set so that segments issued from that node (or very close to it) will have a very low speed allowed but still greater than 0, i.e. for some positive constant $a > 0$ (e.g. $a = 0.05$).

It may be possible for the user to set the reference point as a circular area within the hotspot area (that may or may not contain vertices and/or edges) (Figure 4). If a segment has one of its vertices lying in that circular area, its distance is considered as 0.0. Otherwise, it is the distance between the segment and the circular area boundary.

2.2 Differentiation Between Source and Sink Hotspots

The idea behind this first enhancement is that no traffic rush hour behaves the same in the morning and in the afternoon. If traffic comes in in the morning, it will likely go out in the afternoon. To reflect this, segment edges that are incoming into a source hotspot area, and segment edges that are outgoing from a sink hotspot will not be affected.

A quick way to implement this enhancement is to consider the orientation of the segment edges. Since the edges may represent road segments with curves, the orientation considered will be that going from the origin node to the destination node according to a straight line. Distances will be computed in the same way as before.

Segment edges whose orientation is close to that of the radial vectors to/from the hotspot reference point can easily be associated to an incoming/outgoing edge. Segment edges that are close to perpendicular to a radial vector to/from the hotspot reference point may be considered as outgoing as well as incoming edges. Thus they will be affected by the hotspot in both directions. A threshold will be applied to discriminate limit cases. This threshold corresponds to the maximum angle allowed for the orientation of a segment edge to consider it both as an incoming and outgoing edge.

Instead of actually computing orientation angles, a quicker method is to compare the distances between both the origin and destination of an edge with the hotspot reference point. If both distances are equal (or sufficiently close), this means that the arc forms the base of a quasi-isosceles triangle whose summit is the hotspot reference point. Thus, it can be considered both as an incoming and outgoing edge. The threshold will be set as a value representing the ratio of the shortest of the two distances to the longer one. For example, $t = 0.90$ means that if the smallest distance is at least 90% of the longer one, then the segment edge(s) (either one or 2) will be considered both as outgoing and incoming edges. Hence, they will be affected even if the hotspot is a source or a sink. If a segment is such that the ratio of the distances is smaller than the threshold then, if the segment has two edges (opposite directions), one will be incoming and the other outgoing. If the segment is a one-way edge, then it is either an incoming or an outgoing edge, depending on which of its origin or destination vertex is closest to the hotspot reference point (Figures 5 and 6 provide an example of this).

In Figures 5 and 6 , vertices 5 and 13 are approximately at the same distance from the hotspot reference point. So edges $[5-13]$ and $[13-5]$ are both considered as incoming and outgoing. Consequently, both edges will be affected by the hotspot whether it is a source or sink hotspot. If the hotspot reference is a circular area then any vertices lying within that area and that are connected by some edges will have these edges also considered as both incoming and outgoing and will be affected by the hotspot, whether it is a source or sink hotspot.

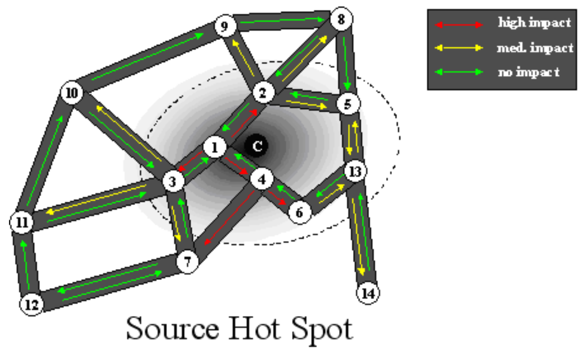

Figure 5: *Example of a Source Hotspot*

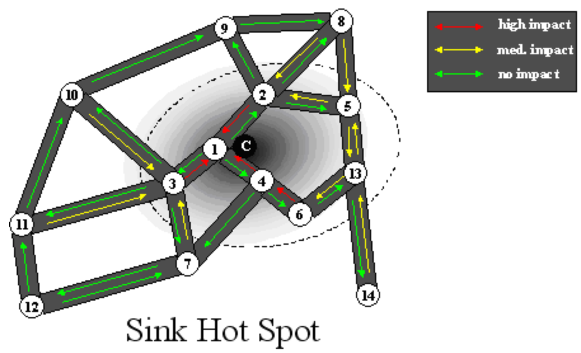

Figure 6: *Example of Sink Hotspot*

2.3 Effect Variation Over Time

Hotspots may affect traffic in a recurrent manner, i.e. at certain periods during a day or week. Moreover, the impact will often be gradually increasing/decreasing during that period of time. A typical example is government offices buildings, where civil servants will gradually arrive in the morning according to an increasing rate until the usual work start time, after which this rate will quickly decrease. The situation is opposite when they are leaving at the end of the day. The exit rate quickly rises at the usual end of work time and slowly decreases as there are fewer employees departing. The first case corresponds to a sink hotspot and the second, to a source hotspot. However, in both cases, the impact on traffic speed is not only dependent on the distance from the reference point but also on the time at which the impact of the additional incoming/outgoing flow occurs.

A mollifying function can be used to dampen the normal distribution over a period of time. The typical behavior of a source hotspot is to have an outgoing flow which quickly rises and then slowly declines over time. On the other hand, the typical behavior of a sink hotspot is to have an incoming flow which slowly increases to a maximum then quickly declines over time. The lognormal, Weibull and Rayleigh distribution functions can be used to represent such a behavior (with suitable parameters – see Annex A). The Weibull distribution has degenerate cases for some values of its parameters so we should avoid it. The Rayleigh distribution is a special case of the Weibull distribution. It has only one parameter, which simultaneously characterizes its shape and scale, so it is less flexible. The lognormal distribution has a shape/scale parameter and a location parameter which allow adjusting its mean, so we propose to use the lognormal distribution.
The default parameters will be set to $\mu = 0$ and $\sigma^2 = 1$

$$(X \sim LogN(\mu, \sigma^2) \longleftrightarrow Y = ln(X) \sim N(\mu, \sigma^2))$$

The user may select different values for μ and σ^2. However, parameters should be selected so as to avoid degenerate cases (e.g. $\mu = 0$ and $\sigma^2 \geq 10$). The variable X represents the time, within the interval of time for which the hotspot has an influence. Since this interval is finite and the lognormal distribution is infinite, the lognormal will be truncated at 95% of its cumulative distribution function and normalized over the length of the time interval for the hotspot. The lognormal distribution will be used for source hotspots. For sink hotspots, the lognormal will be applied to the complement of X in the time interval of the hotspot (see below); i.e. if D is the length of the hotspot time interval, then:

- $X \sim LogN(0,1), 0 < a \leq x \leq D$ if the hotspot is a source

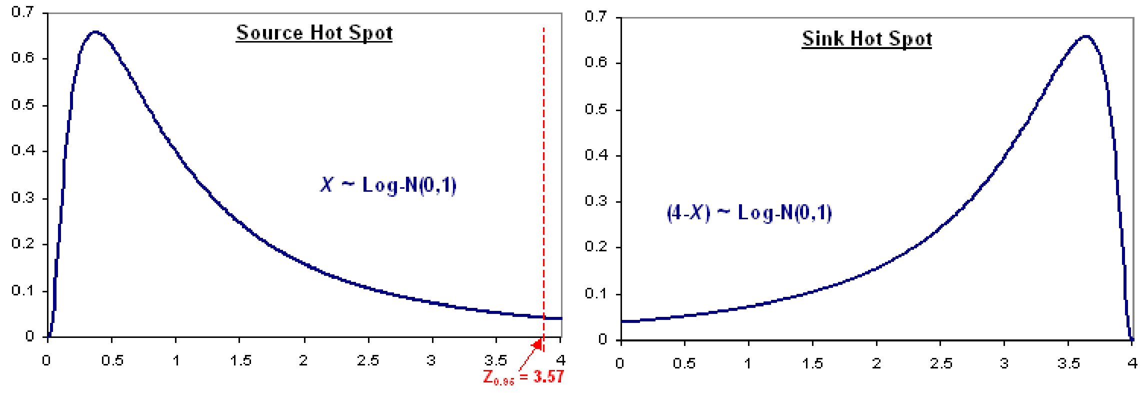

Figure 7: *LogN Distributions for Source and Sink Hotspots*

- $(D-X) \sim LogN(0,1), 0 \leq x \leq D$ if the hotspot is a sink

The values of the lognormal density will be normalized over the mode of the distribution, so that the maximum will correspond to 1. For any value of X, the normalized lognormal density value will represent the proportion by which the normal density values for the distance impact in the hotspot will be multiplied. At the peak time in the time interval (peak of the lognormal distribution), the normal distribution selected will be applied without modifications while it will be dampened for values lower or greater proportionally to the lognormal curve.

3 Path Search Based Solution

The use of the normal distribution on the distances between the segment edges and the hotspot reference point does not correctly process all situations. For instance, a network may be such that within a source hotspot, a one-way segment goes in the direction of the reference point just to connect with another that goes away from it. It might thus be the case that an outgoing vehicle may have to take a segment arc that brings it closer to the reference point for some distance. With the implementation described in the previous section, this edge would be considered as an incoming arc and thus be unaffected by the hotspot, while it could be affected if it is the only way to reach the other outgoing segment or if it is taken by many vehicles.

Figure 8 illustrates a situation where vehicles that need to go from vertex 4 to vertex 2 have to go through edges $[7-3]$ and $[3-1]$ which are considered as incoming edges and thus would be unaffected by the source hotspot while they may actually have their traffic speed affected by outgoing traffic.

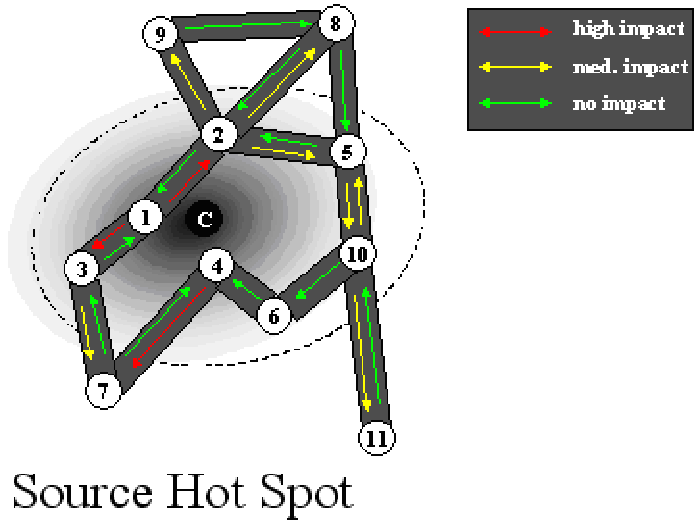

Figure 8: *Example of Network with Multiple Direction Segments*

3.1 Iterative Shortest Path Searches

The use of shortest paths to establish the traffic speed impact on each segment edge instead of the normal distribution. In some situations, outgoing/incoming vehicles may have to take a segment edge that brings them closer/farther to/from a source/sink hotspot reference point (because this is the only way or it is a shortest path). A majority of vehicles will usually take the shortest path to get into or out of a hotspot.

The basic idea is to extract the subgraph vertices and edges that are part of the hotspot. If the reference point is not located at a vertex, set it to the closest vertex that has at least one outgoing/incoming edge. For each other vertex in the subgraph, compute its shortest path to/from the node used as reference point using a shortest path algorithm (depending if the hotspot is a source or sink the shortest path between a vertex and the reference point may differ). For any vertex that does not have a path to/from the vertex used as a reference point, remove that vertex and all its adjacent edges from the subgraph. For each remaining edge, count 1 each time this edge appears in one of the shortest paths computed. Each edge is assigned an impact on its maximum traffic speed directly proportional to its corresponding count tallied over all shortest paths. The impact values are normalized over the larger count to yield an impact value between 0 and 1 (set a minimum threshold value if required). Apply the impact as a proportion of the maximum speed allowed as before.

The impact values computed may be dampened in exactly the same way as described in 2.3, except that the normalized lognormal values will be applied to the impact computed from the counts and not from a normal distribution with regard to distance.

This approach has the advantage that impossible paths or segments that lie within the hotspot but cannot be reached (due to one-ways or because they are disconnected) will not be affected by the hotspot. This approach uses the fact that recurrent traffic vehicles usually know the shortest path and will usually use it to arrive to destination.

In some situations, other behaviors can be observed. For example, if an area has major boulevards or speedways with many lanes, it can be the case that, though there could be a shortest way using smaller secondary streets, many vehicles will prefer to take the main arteries. In the presence of such information, the previous scheme could be applied by giving supplementary count values to such major arteries.

This final approach encapsulates the generality of the previous ones, but adds an analytical aspect that makes it more likely to better reflect traffic behavior. The flow of traffic is still represented generally with a global area of effect. A center of attraction is still defined. Yet, motorists' behavior is reflected by emulating the path finding behavior each of them is more likely to have. Let us now go over each approach in a more detailed fashion.

4 Algorithms

Here is an overview of procedures that can be used to implement the various approaches described before.

4.1 Basic Solution

Algorithm 1 describes the steps necessary to obtain the maximum traffic speed (MTS), for any edge present in a hotspot.

4.2 Differentiation Between Source and Sink Hotspots

Algorithm 2 describes the steps required to obtain the maximum traffic speed for any edge, using algorithm 1. In this case, we take into account whether the hotspot is a sink or source.

Algorithm 1 Compute Basic Traffic

let $supN$ be the $N(\mu_N, \sigma_N^2)$ normal density at mode value $(mode = \mu_N)$
for all segments in the subgraph regardless of their number of edges **do**
  if the segment has both its adjacent vertices in the hotspot **then**
    find the adjacent vertex n which has the smallest distance d to the reference point or circular area
  else if segment has only one of its adjacent vertices in the hotspot **then**
    let that node be n and d its distance to the reference point or circular area
  end if
  let d' be the length of the director radius from the reference point or area passing through n and intersecting the hotspot boundary
  let $x = d/d'$ the ratio of distance to director radius
  let a be the distance ratio threshold
  if $x < a$ **then**
    let $w = a$
  else
    let $x' = x * 1.96$ {normalize wrto $CDF = 0.95$ of $STND(0,1)$}
    let $y = x' * \sigma_N + \mu_N$
    let p be the $N(\mu, \sigma^2)$ normal density at y
    let $w = 1 - p/supN$ {$p/supN$ is the normalization of p over the mode density}
  end if
  for all edge in the segment **do**
    set maximum traffic speed $MTS(edge) = w * MSA(edge)${MSA = Maximum Speed Allowed}
  end for
end for

4.3 Effect Variation Over Time

Algorithm 3 describes the steps necessary to make the maximum traffic speed computed with algorithms 1 and 2 vary over time.

4.4 Iterative Shortest Path Searches

Algorithm 4 details the necessary steps to compute the maximum traffic speed with the path search approach.

Algorithm 2 Compute Sink/Source Traffic

let $supN$ be the $N(\mu_N, \sigma_N^2)$ normal density at mode value $(mode = \mu_N)$
for all segments in the subgraph regardless of their number of edges **do**
  Select the segment two adjacent vertices n_1 and n_2
  Compute the distances d_1 and d_2 between each vertex and the reference point or circular area
  if $max(d_1, d_2) = 0$ **then**
    let $k = 1$ {both n_1 and n_2 are inside the reference circular area}
  else
    let $k = min(d_1, d_2)/max(d_1, d_2)$
  end if
  if $k < t${Where $0 < t < 1$ is the threshold for the ratio of distances} **then**
    if hotspot is a source **then**
      if $d_1 < d_2$ **then**
        if the segment has an edge going from n_2 to n_1 then remove that edge from the subgraph
      else
        if the segment has an edge going from n_1 to n_2 then remove that edge from the subgraph
      end if
    else
      if $d_1 < d_2$ **then**
        if the segment has an edge going from n_1 to n_2 then remove that edge from the subgraph
      else
        if the segment has an edge going from n_2 to $n_1$then remove that edge from the subgraph
      end if
    end if
  end if
  if the segment has at least one edge still in the subgraph **then**
    apply one iteration of algorithm 1 to the segment
  end if
end for

Algorithm 3 Compute Traffic Variation Over Time

let $[t_D, t_F]$ denote the interval of time during which the hotpot has an influence
let $t \in [t_D, t_F]$ be the time at which the impact is to be evaluated {if $t < t_D$ or $t > t_F$ then the hotspot has no impact on any segment edge}
let $supN$ be the $N(\mu_N, \sigma_N^2)$ normal density at mode value $(mode = \mu_N)$
let $supLogN$ be the $LogN(\mu_{LogN}, \sigma_{LogN}^2)$ lognormal density mode value $(mode = e^{\mu_{LogN} - \sigma_{LogN}^2})$
for all segments in the subgraph regardless of their number of edges **do**
  apply one iteration of algorithm 2
  if the segment has at least one edge still in the subgraph **then**
    for all edge in the segment **do**
      let w be the speed reduction factor on that edge as computed by the iteration of algorithm 1 (within the iteration of algorithm 2)
      if hotspot is a source **then**
        let $t' = (t - t_D)/(t_F - t_D), 0 < t' < 1$ {translate and normalize over hotspot time interval}
      else
        let $t' = (t_F - t)/(t_F - t_D), 0 < t' < 1)$ {complement, translate and normalize over hotspot time interval }
      end if
      let $CDF_{0.95} = e^{((1.15\sigma_{LogN}\sqrt{2}) + \mu_{LogN})}$ {inverse $ERF(x)$ for lognormal $CDF = 0.95$}
      let $x = t' * CDF_{0.95}$ {normalize wrto lognormal $CDF = 0.95$}
      let p be the $LogN(\mu_{LogN}, \sigma_{LogN}^2)$ lognormal density at x
      set $w' = 1 - p/supLogN$ {$p/supLogN$=normalization of p over mode density}
      if $w' * w < a$ {where a is the threshold value for impact wrto distance} **then**
        set the maximum traffic speed on the segment edge of the subgraph as $MTS(edge) = a * MSA(edge)$
      else
        set the maximum traffic speed on the segment edge of the subgraph as $MTS(edge) = w' * w * MSA(edge)$
      end if
    end for
  end if
end for

Algorithm 4 Compute Shortest-Path-Based Traffic

if the hotspot reference location is a point which is a not a vertex in the network **then**
  set the reference point to the location of the closestvertex which has at least one outgoing/incoming edge according to whether the hotspot is a source/sink
end if
if the hotspot reference location is a circular area that does not enclose any network vertex **then**
  center it on the nearest vertex which has at least one outgoing/incoming edge according to whether the hotspot is a source/sink (the hotspot reference area may include more than one vertex)
end if
for all vertices in the subgraph not the reference vertex **do**
  compute the shortest path to the reference point of the hotspot
  if the hotspot is a source **then**
    outgoing paths are computed
  else
    incoming paths are computed
  end if
end for
if hotspot reference is a circular area containing more than one node **then**
  repeat the computations of shortest paths for every node in the hotspot reference area
end if
remove all vertices in the subgraph that do not appear in at least one of the shortest paths computed and remove their adjacent edges from the subgraph
for all edges e_i remaining in the subgraph **do**
  let $p(e_i) > 0$ be a weight value assigned to edge e_i (this weight will be cumulated with the frequency of the edge in all the shortest paths), $p(e_i)$ may be constant or variable for every edge of the subgraph
  let $c(e_i)$ be the number of times edge e_i is encountered in all the shortest paths computed
  let $w(e_i) = 1 - \frac{c(e_i)*p(e_i)}{max_i[c(e_i)*p(e_i)]}$
  if $w(e_i) < a$ **then**
    set $w(ei) = a$ {where a is the threshold value set for impact as before}
  end if
  set the maximum traffic speed on arc e_i as $MTS(e_i) = w(e_i) * MSA(e_i)$
end for
if hotspot influence is to be variable over a period of time **then**
  apply algorithm 3 for every edge of the subgraph using w as computed above
end if

5 Conclusion

In order to provide path planning abilities in dynamic urban environments, it is essential to take traffic into account. This document presents three novel ways of efficiently acquiring preliminary traffic information. While the three approaches have differences, they are all based on a very instinctive concept of traffic attraction points (hotspots). Such points attract people at different moments of the day, directly impacting the flow of traffic.

The first approach applies a distribution function that affects the traffic congestion model within the hotspot. This first approach provides a computationally simple solution to the problem. While lacking the subtlety of the two following techniques, it represents a good starting point, allowing for a quick way to implement the hotspot concept. It is possible to add to the first approach, slowly building towards the second proposed solution. A first possible improvement is to include the difference between sink and source hotspots. It is obvious that if people go towards a hotspot at a particular point in the day, they will eventually leave it. Sink hotspots represent attraction points that people head towards, and source hotspots represent attraction points that people depart from. Following from that is the notion that motorists may not arrive at an attraction point at a steady, linear pace. For instance, if we consider a big office building, a few people may start arriving slowly from six in the morning, but a lot more will come in at eight or nine. This concept of effect variation over time was also described and added to the solution. The second proposed solution is more complex, flexible and more likely to better model reality. The last approach takes into account the behaviour of motorists that have a certain knowledge of the road system. As they grow more accustomed to a particular part of the city, drivers will start avoiding certain areas that they know to be slower. Each driver tries to find an optimal path to get to and from a hotspot. This is the idea reflected in the last approach. Iterative shortest paths are computed to establish the traffic impact on each segment. Generally, the more a road segment is likely to be part of an optimal path going from or to a hotspot, the higher impact the hotspot will have on it.

The distribution based and shortest path based solutions each represent a different take on the same problem. Both add the hotspot effect to the traffic model. Using both approaches against a real road network will allow to better identify in which particular cases each approach is more likely to provide a better initial picture. It is possible that the use of both approaches might yield an initial model closer to real world data. The construction of a hybrid solution, merging results from both approaches, may have to be considered.

References

1. Pigeon, L., Lecornu, L., Bergeron Guyard, A., and Bergeron, A. (2005). Navigation in Extreme Environments - Scipio Optipath. *DRDC-Valcartier TR 2005-529*.

2. Pigeon, L., Burt, G., and Riffou, F. (2005). SCIPIO Urban Suite, Campaign plan 2005. *DRDC-Valcartier TM 2005-530*.

3. Nakata, T. and Takeuchi, J.i. (2004). Mining traffic data from probe-car system for travel time prediction.. In *KDD (Conference on Knowledge Discovery and Data Mining)*, pp. 817–822.

4. Chen, M. and Chien, S.I.J. (2002). Dynamic Freeway Travel Time Prediction Using Probe Vehicle Data: Link-based vs. Path-based. *Journal of the Transportation Research Record*, (1768), 157–161.

5. Chien, S.I.J. and Kuchipudi, C. Mouly (2003). Dynamic Travel Time Prediction with Real-Time and Historic Data. *Journal of Transportation Engineering*, **129**(6), 608–616.

This page intentionally left blank.

Annex A
Probability Distributions

A.1 Probability Distribution Function of a normal random variable $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

parameters:
$\mu =$location (=mean)
$\sigma =$scale (=standard deviation)

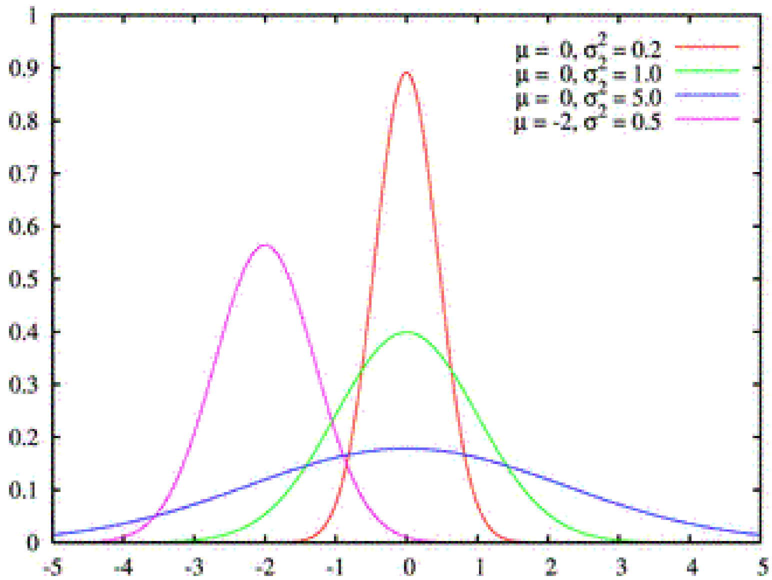

Figure A.1: *Normal Distribution*

A.2 Probability Distribution Function of a lognormal random variable $X \sim Log - N(\mu, \sigma^2)$

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{(ln(x)-\mu)^2}{2\sigma^2}}, x > 0$$

parameters:
$-\infty < \mu < \infty$, location parameter (=mean of $ln(x)$)

$\sigma > 0$, shape/scale parameter(=standard deviation of $ln(x)$)

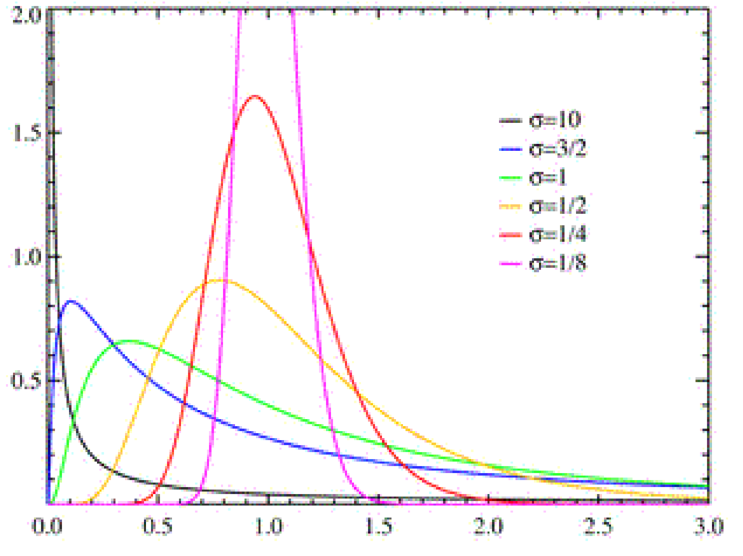

Figure A.2: *Lognormal Distribution ($\mu = 0$)*

A.3 Probability Distribution Function of a Weibull random variable $X \sim Weibull(k, \lambda)$

$$f(x) = (k/\lambda)(x/\lambda)^{k-1}e^{-(x/\lambda)^k}.x > 0$$

parameters:
$k > 0$, shape parameter
$\lambda > 0$ scale parameter

A.4 Probability Distribution Function of a Rayleigh random variable $X \sim R(\sigma^2)$

$$f(x) = \frac{x}{\sigma^2}e^{-x^2/2\sigma^2}, -\infty < x < \infty$$

parameters:
$\sigma > 0$, shape/scale parameter

DRDC Valcartier TM 2006-792

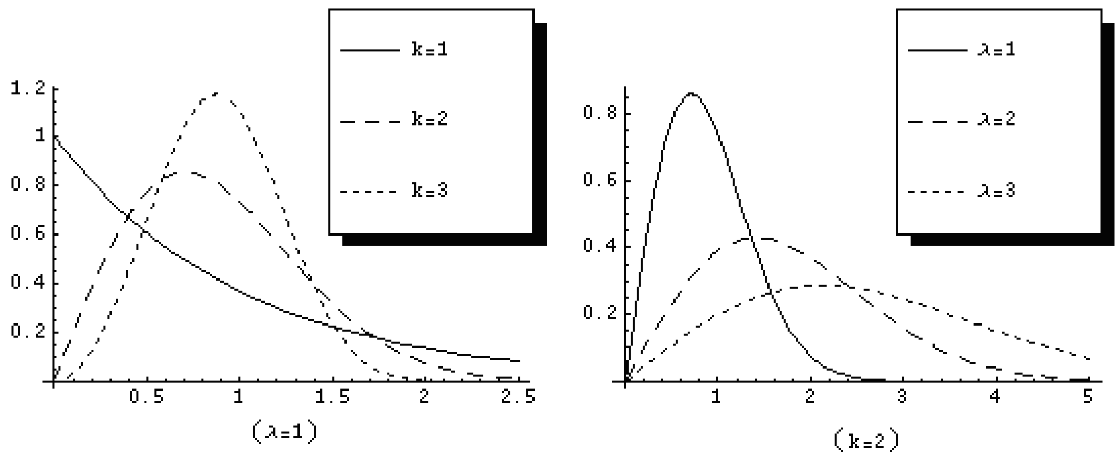

Figure A.3: *Weibull Distribution*

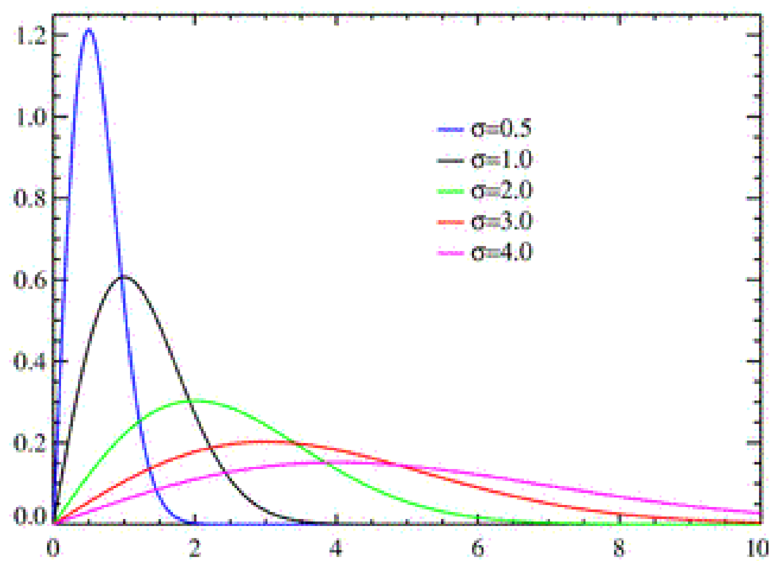

Figure A.4: *Rayleigh Distribution*

This page intentionally left blank.

Distribution list

Internal distribution

1 Director General

3 Document library

1 Head/RI

1 Head/SAD

1 Head/SDS

1 Mr. Alexandre Bergeron Guyard (author)

1 Dr. Luc Pigeon (author)

1 Mr. Alain Bergeron

1 Mr. Éric Dorion

1 Mr. Alain Bouchard

1 Dr. Alain Auger

1 Mrs. Régine Lecocq

1 Mr. Jean Roy

1 Dr. Pierre Valin

Total internal copies: 16

External distribution

1 DRDKIM (PDF)

NDHQ - 101 Colonel By Drive, Ottawa, ON, K1A 0K2

1 Mr. Yves van Chestein

DRDC HeadQuarters - 305 Rideau Street, Ottawa, ON, K1A 0K2

1 Major Shawn Hoag

CFB Kingston - PO Box 17000, Station Forces, Kingston, ON K7K 7B4

1 Dr. Jerôme Lévesque

CFB Valcartier - PO Box 1000, Station Forces, Courcelette, Qc, G0A 4Z0

1 Major Benoit Côté

Thales Canada, Quebec Business Unit - 3101 Avenue Des Hôtels, Bureau 18D, Sainte-Foy, QC, G1W 4W5

1 René Proulx

Total external copies: 6

Total copies: 22

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) Defence R & D Canada – Valcartier 2459 Pie-XI Blvd North, Québec, QC, Canada, G3J 1X5	2. SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable). UNCLASSIFIED

3\. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title).

Production of preliminary traffic data for path planning in urban environment: From probability distributions to iterative shortest path algorithms

4\. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

Guyard, A. B. ; Pigeon, L. ; Proulx, R.

5. DATE OF PUBLICATION (month and year of publication of document) March 2008	6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc). 36	6b. NO. OF REFS (total cited in document) 5

7\. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered).

Technical Memorandum

8\. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include address).

Defence R & D Canada – Valcartier
2459 Pie-XI Blvd North, Québec, QC, Canada, G3J 1X5

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Specify whether project or grant). 1430jm12pm01	9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written).
10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique.) DRDC Valcartier TM 2006-792	10b. OTHER DOCUMENT NOs. (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11\. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Defence departments and defence contractors; further distribution only as approved
( ) Defence departments and Canadian defence contractors; further distribution only as approved
( ) Government departments and agencies; further distribution only as approved
( ) Defence departments; further distribution only as approved
( ) Other (please specify):

12\. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution beyond the audience specified in (11) is possible, a wider announcement audience may be selected).

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

The SCIPIO Optipath module aims at providing path planning abilities in evolving urban environments. To effectively do so, the system has to take into account numerous factors (urban environment, sensor/humint force tracking, threats, etc.). Among those factors, traffic is an essential one. Traffic has a direct impact on how easy it is to maneuver in a city. However, traffic data can be hard to acquire. While it is possible to obtain maps of a city and to track enemy forces or threats, traffic data is very hard to collect. Attempts have been made using numerous probe cars over long periods of time (weeks or months). Such constraints cannot easily be dealt with in a deployment context. We therefore, must have a fast and efficient way to provide some usable preliminary traffic data. This paper presents a way to obtain such data. It is based on high-density attraction points (or hotspots), which dictate traffic behavior patterns.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title).

SCIPIO
Traffic
Optipath
Urban Operation
Urban Environment
Theatre Modeling

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE DÉFENSE

www.drdc-rddc.gc.ca