Emitter-Platform Association

Problem presented by

Neil Cade and Neil Brearley

Selex Galileo Ltd

Executive Summary

Intercepted RF electromagnetic signals provide a good long-ranged source of information on the motions and activities of people, vehicles, installations and organisations. For those emissions that are detected, traditional tracking methods are used to associate the separate low level interceptions and average their characteristics to obtain tracks of the source location and characteristic patterns of the emissions. The Study Group was asked to provide a prediction of the number of underlying source platforms and the association between the emissions and platforms.

Version 3.0
20 July 2010

Report author

Tristram Armour (Industrial Mathematics KTN)

Contributors

Tristram Armour (Industrial Mathematics KTN)
Misaina Andrianjafinandrasana (University of Witwatersrand)
David Barton (University of Bristol)
John Billingham (University of Nottingham)
David Sibley (University of Bath)
Florian Theil (University of Warwick)
Nicholai Peremezhney (University of Warwick)
Switalski Przemyslaw (Uniwersytet Warszawski)
Tom Ranner (University of Warwick)

ESGI73 was jointly organised by
The University of Warwick
The Knowledge Transfer Network for Industrial Mathematics

and was supported by
The Engineering and Physical Sciences Research Council

Contents

1 Introduction

1.1 Problem description

(1.1.1) Intercepted RF electromagnetic signals provide a good long-ranged source of information on the motions and activities of people, vehicles, installations and organisations. Such sources range from mobile phones at the low frequency end, though surveillance radar (air traffic control) to millimetre guidance radar (car collision avoidance), all of which produce intermittent pulse signals of varying frequency, inter-pulse timing and pulse shape. This is a very broad spectrum and typically, broadband RF receivers will potentially be able to detect several hundreds of sources at any time instant. For those emissions that are detected, traditional tracking methods are used to associate the separate low level interceptions and average their characteristics to obtain tracks of the source location and characteristic patterns of the emissions.

(1.1.2) This traditional tracking problem is complicated for several reasons:

- The emissions are sporadic, consisting of short bursts of emission interspersed with long quiescent intervals.
- The individual sources have multiple modes of operation (e.g. mobile phones may be transmitting voice or data station-polling signals).
- Location information is available in terms of noisy measurements of azimuth, elevation angle and range.
- Platforms can have multiple sources of emission (e.g. a ship may have several different types of Radar, or a bus may have several people using their mobile phones at the same time).
- The interceptor sensors are also likely to be on moving platforms and will not necessarily have a consistent visibility of the sources (occlusion, multi-path, etc).
- The emission is dense enough that the emission patterns from different sources are bound to overlap with at least one other source.

Therefore it is inevitable that the traditional tracking will have introduced some additional track errors from mis-associations that in turn result in incorrect classifications and location distortions. Given intercepted Radio Frequency (RF) emissions, the Study Group was asked to provide a prediction of the number of underlying source platforms and the association between the emissions and platforms.

(1.1.3) The proposed study took this emitter track data from the intercepted RF emissions as given. The multi-target track data is a sequence of time-stamped state vectors comprising continuous (angle) components, estimates of the maximum possible ranges of the data sources, an identity that associates the individual emission belonging to the separate tracks and the associated uncertainties of these characteristics. A simulation was written to create this data.

(1.1.4) The task was to provide a best many to many match, supported by some measure of the quality of match, between the emissions and potential platforms at all times, on the basis of previously seen data. The issues that needed to be addressed include:

- The emission sequences associated with a single identity may be wrong. For example the same type of emission might come from multiple platforms of the same type and may therefore have been incorrectly associated with a single track.
- Platforms of the same type can have very different emissions.
- Platforms can have emissions overlapping in time.
- Ambiguities may become resolved as the targets approach the sensor system or as different platforms move relative to each other.
- The accuracy of the track data can vary greatly between different tracks and over the evolution history of a single track.
- The need to avoid discontinuous jumps in the mappings as time evolves. Ultimately, the primary interest is in the underlying platforms and it is particularly disconcerting if the solution chatters between almost equally likely alternatives.

1.2 Assumptions

(1.2.1) Due to the complexity of the original problem it was useful to make the following assumptions agreed by the industrialist. However, the final algorithm that we put forward in this report does not require the second assumption to be true.

(1.2.2) **Assumption 1**: Each track detected comes from a single platform. In the original description, the tracking software can make mistakes and start tracking a different platform, but it was agreed that these cases are rare enough that we can ignore it for the duration of the Study Group.

(1.2.3) **Assumption 2**: A platform can produce at most one track at any time. In reality, platforms can produce more than one track at a time but it is rare enough for it to be a reasonable assumption.

2 Problem Formulation

2.1 Definitions

(2.1.1) Let $n(t)$ be the number of tracks in the time interval, $[0, t]$.

(2.1.2) Each track i at time t, $i \leq n(t)$, can be defined by the collection, $\{r_i, a_i, \tau_i, \alpha_i\}$ where r_i is the set of range values, a_i is the set of azimuthal values, τ_i is the set of timestamps for the track data and α_i is the type of the transmission.

(2.1.3) Track i overlaps with track j if $\tau_i \cap \tau_j$ is non-empty. Note that under assumption 2, this implies that two different platforms created these tracks.

2.2 Problem breakdown

(2.2.1) The problem can be split into four parts:

(a) Find the trajectory of each track
(b) Compare the trajectory information for each track
(c) Find the number of likely platforms and assign tracks to platforms
(d) Hysteresis for the decision-making

(2.2.2) The challenge at the Study Group was to develop an algorithm or framework that fused discrete, platform specific information and continuous geographical information about the track data to make the best assignment of the unknown platforms with the tracks.

2.3 Simulation

(2.3.1) The simulation is the forward model of the problem with features prespecified by the industrialist. We can check the accuracy and robustness of our solution to the inverse problem by adding noise to the signal. Accuracy is subject to how well the simulation reflects the real problem, therefore, it is important to simulate the data with the same properties of the collected data. This section describes the simulation of data used in this project.

(2.3.2) At each instant of the simulation, the idealised signal gives the position of the platform to our receiver. We assume the platform moves with constant velocity through time. More precisely, $x_0(t) = x_0 + \cos(\theta)\, t$ and $y_0(t) = y_0 + \sin(\theta)\, t > 0$, for some arbitrary angle θ. We set the final time to be the time it takes the platform to exit the unit circle $t_m = 2\cos(3\pi/2 - \theta)$. We can derive the true azimuth $a_0(t)$ and range $r_0(t)$ from the position at time t by changing to polar coordinates.

(2.3.3) The azimuth received by our platform, $a(t)$, is simulated by adding noise to the true azimuth $a_0(t)$. That is, $a(t) = a_0(t) + W_t$, where the noise $W_t \sim N(0, (\pi/36)^2)$ is normally distributed with zero mean and $\pi/36$ standard deviation. The received range is obtained by scaling the true range $r_0(t)$ by a random factor. We use, $r(t) = r_0(t) \cdot 2^{U_t}$, where $U_t \sim U(-4/6, 1/6)$ uniformly distributed with parameters $(-4/6, 1/6)$, as per request of the industry representative.

(2.3.4) Finally, the durations of the transmission periods and the quiet periods are exponentially distributed with parameter, 0.1. That is, the start and end points of the transmission together is a Poisson process with rate 0.1. The corresponding position $(x(t), y(t))$, if needed, can be derived from $(r(t), a(t))$. Figure 1 depicts the simulated azimuth and range. Figure2 depicts the corresponding position time series.

(2.3.5) Note that, because $r(t)$ has a large error margin, setting $x(t) = r(t)\cos(a(t)$ and $y(t) = r(t)\sin(a(t))$ will have large error margins.

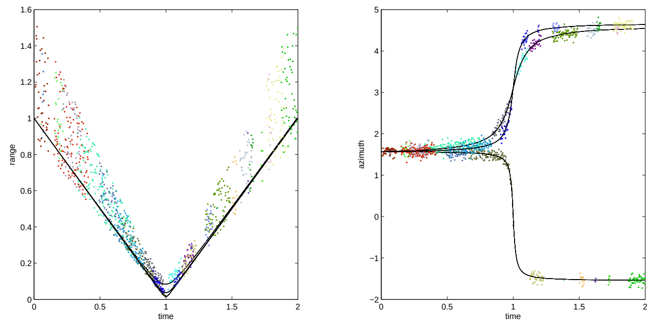

Figure 1: Actual and noisy azimuth and range from three platforms. Colors represent different tracks.

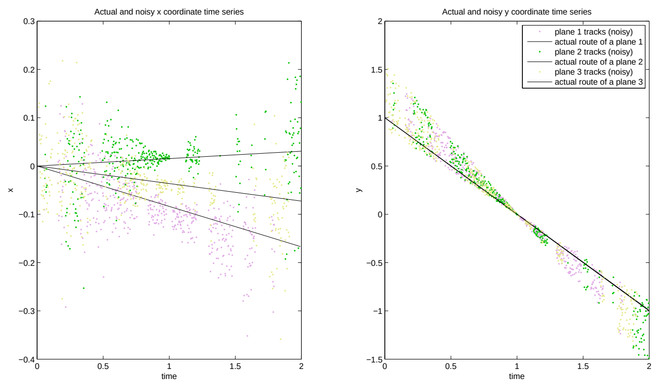

Figure 2: Actual and noisy x and y positions from three platforms. Colors represent different tracks.

3 Assigning platforms to tracks

3.1 Graph representation of the problem

(3.1.1) Our problem can be represented using a graph, call it, $G = (V, E, W)$ where V is the set of nodes, E is the set of edges and W is the set of weights associated with each edge of the graph. Let each track in the problem be represented by a node and let t_i be represented by node i.

(3.1.2) If there is an edge between node i and node j, this means that the tracks could have been made by the same platform. The likelihood that the tracks were made by the same platform is represented by the weights $w_{ij} = w_{ji}$ where each weight is in the interval, $[0, 1]$. The weights represent how likely the two tracks were created by the same platform.

(3.1.3) Using assumption 2 in (1.2), we can start to create a useful lower bound for the number of platforms in the system at time t by counting the number of overlaps in the tracks. If there are m overlaps at time t, we can conclude that there are at least m platforms. If track i overlaps with track j, there is no edge between nodes i and j.

(3.1.4) There are a number of ways we can assign the weights in W and much depends on the prior assumptions we make about the movement and the activities of platforms we expect to capture using the tracker. Due to time constraints, the Study Group participants set $w_{ij} = 1$ for all tracks i and j that overlapped.

3.2 Colouring the complement

(3.2.1) Let the complement of the graph, G be defined as $\hat{G}$. In the complement graph, if an edge exists between node i and node j, this implies that there is an overlap between tracks i and track j. Each graph colouring of $\hat{G}$ is a permissible assignment of platforms to the tracks, where if two nodes are coloured the same colour, they come from the same platform.

(3.2.2) Note that, there is more often than not, more than one colouring of the graph. A new graph, call it G', is created when we remove the edges from G that join nodes of different colour. Let C be the set of edges in G and in G'. Let C^c be the set of edges in G but not in G'. We calculate the fitness of the track allocation by $f_C = \prod_{(i,j)\in C} w_{ij} \prod_{(i,j)\in C^c}(1 - w_{ij})$. The colouring that minimises f_C is our best platform allocation.

(3.2.3) Let us describe the algorithm using the example seen in Figure 3. These tracks can be used to create the graph, G, which can be seen as the graph in (A) in Figure 4. The graph labelled (B) is $\hat{G}$ and the shows the possible colourings of each node. The graph in (C) is G'.

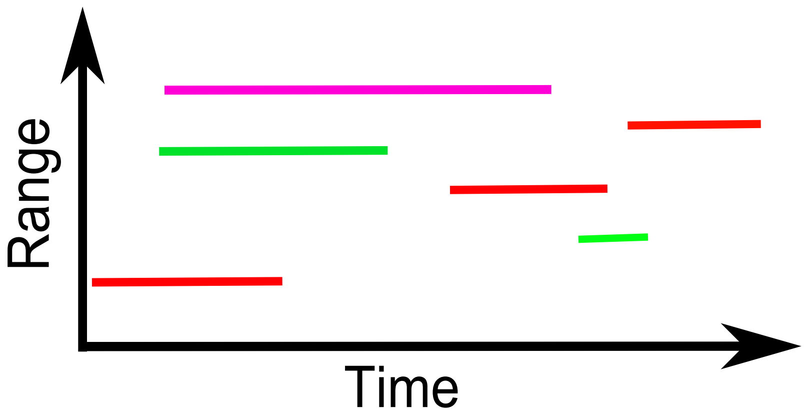

Figure 3: Idealised track information

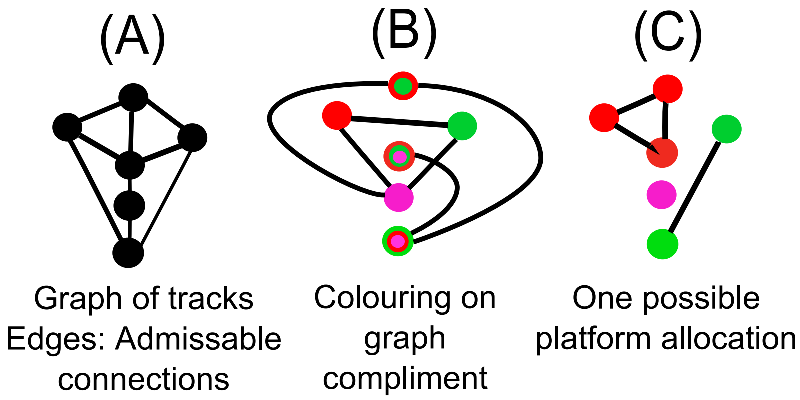

Figure 4: Colouring the graph

(3.2.4) The fitness function here was not used at the Study Group due to the weight assignments. By comparing the colouring fit and the true colouring (obtained by looking at the simulated data) we derived a separate fitness function.

(3.2.5) Note that platforms correspond to completely connected subgraphs of G. But, completely connected subgraphs may be made up by more than one platform.

4 Conclusions

4.1 Results

(4.1.1) The Study Group applied this graph theoretic approach on a single case with 5 platforms and approximately 30 tracks. The tracks of the simulation can be seen in Figure 5. We ran the platform assignment algorithm to find the minimum number of platforms required and assign the tracks to the platforms. The algorithm correctly found the number of platforms but the assignment of the tracks to the platforms was very problem dependent.

(4.1.2) Let J be the set of indices for platforms, K_j the set of tracks associated with platform j in a particular colouring. The fitness function of the colouring was calculated by the following formula,

$$f = \sum_{j\in J}\sum_{k\in K}\sum_{t\in\tau_k}\left(\arctan((x_1^j + tv_1^j)/(x_2^j + tv_2^j)) - a_k(t)\right)^2 \tag{1}$$

where x_1^j, x_2^j are the initial positions along the x-axis and y-axis respectively for platform j, v_1^j, v_2^j the initial velocities along the x-axis and y-axis respectively for platform j and $a_k(t)$ is the azimuth at time t for track k.

(4.1.3) Figure 6 shows the best colouring found had a fit value of 0.032 and the track assignment is represented by the top five graphs in the Figure. The bottom five shows the fit of the last colouring attempt out of 10 attempts.

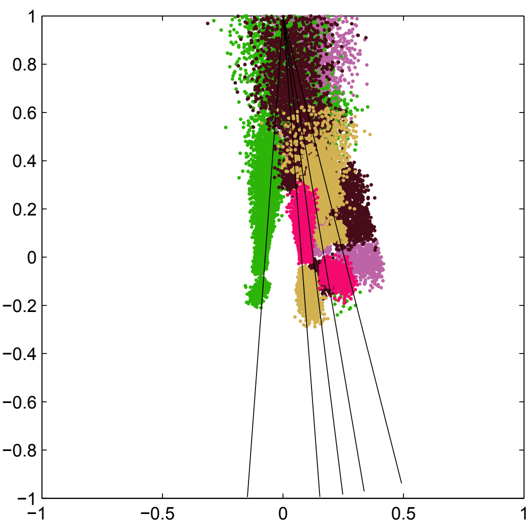

Figure 5: Fitting 5 platforms to the simulated tracks

4.2 Remarks

(4.2.1) The idea behind the graph theoretic approach is to enable the incorporation of probabilistic information (*e.g.*position measurements, known emitter characteristics) and "hard" rules based on track type information and exclusion requirements. The graph is not intended for presentation to the end-user; it is purely a data storage structure.

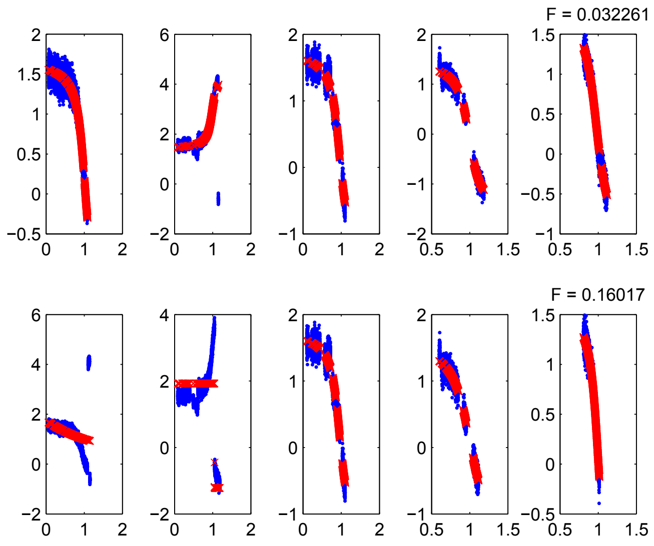

Figure 6: Best colouring and current colouring

(4.2.2) The first attempts in implementing the graph theoretic approach were promising. However, in general, there are many possible colourings of the graph and each colouring is a different assignment of the tracks. Each colouring takes less than one second with 30 tracks but better colouring algorithms will be needed to find an optimal colouring.

(4.2.3) An example framework for the full algorithm of assigning tracks to platforms can be found in the Appendix.

(4.2.4) G is an example of an interval graph and it is well known that these can be coloured by using a greedy algorithm. However, the complement of an interval graph cannot be coloured necessarily in the same way.

4.3 Further work

(4.3.1) The Study Group gave ideas on how to assign the weights in W but did not go as far as developing an algorithm. We suggest that Selex investigate different ways of using the track information to assign the weights in W to be in the interval, $[0, 1]$. We can also incorporate any probabilistic information from knowledge about the types and any other information on the weights in W, possibly by Bayesian methods.

(4.3.2) Randomly colouring the graph can be computationally expensive, therefore, we would advise Selex to develop a colouring algorithm that would improve the fitness of the colouring with respect to the weights, W iteratively.

(4.3.3) Further work is needed on the heuristics as the algorithm as it stands will allow the solution to jump. Using some real data instead of the simulation would allow Selex to see what is required here and how problem specific this is.

5 Appendix and References

5.1 Example framework

(5.1.1) The algorithm described here was not been implemented in its entirety. In contrast to the presented work at the Study Group, this framework is not limited to a single emitter per platform. This framework is untested and would therefore require a full investigation.

(5.1.2) In the graph, each track is a node. Associated with each node is the current position, the last-seen time and any needed signal characteristics. When a track is staled-out, the node is removed from the graph. Edges between nodes indicate that two tracks could be associated to the same platform; the edges are weighted with a probability of association. Whenever data for a new track is obtained a new node is created with edges to the other nodes in the geographical vicinity (also in accordance with known emitter type rules). Whenever data for an existing track is obtained, the weighting probabilities for the edges of that node are updated according to the new information (e.g., new position relative to the other nodes). Should the probability for an edge decrease below a critical threshold, the edge should be removed. New edges should never be added since the lack of an edge indicates that the possibility of a connection was previously ruled out.

(5.1.3) Periodically the graph is frozen in time and copied for use in the platform detection routines. A second threshold can be applied to the copied graph to remove improbable (but not impossible) edges. The problem of platform detection becomes a problem of partitioning the graph into cliques (completely connected subgraphs). (A platform will be a completely connected subgraph, however, a completely connected subgraph may not be a single platform.) It is unlikely that the cliques will be uniquely determined; instead a heuristic algorithm should be used in the clique detection step that takes into account the probabilities on the edges. The present implementation performs an exhaustive search, which may or may not be feasible depending on the data received.

(5.1.4) Additional comments:

- It may be useful to use a Kalman filter with a model for a point mass with a constant velocity for each track to keep a continually updated position estimate for each track.

- Once the platform allocations have been decided the average of the track positions is used for the platform position.

5.2 References

(5.2.1) Allwright, David and Gould, Tim and Gravesen, Jens and Leese, Robert and Petersen, Henrik (2006): Graph colouring for office blocks (Study Group report)