

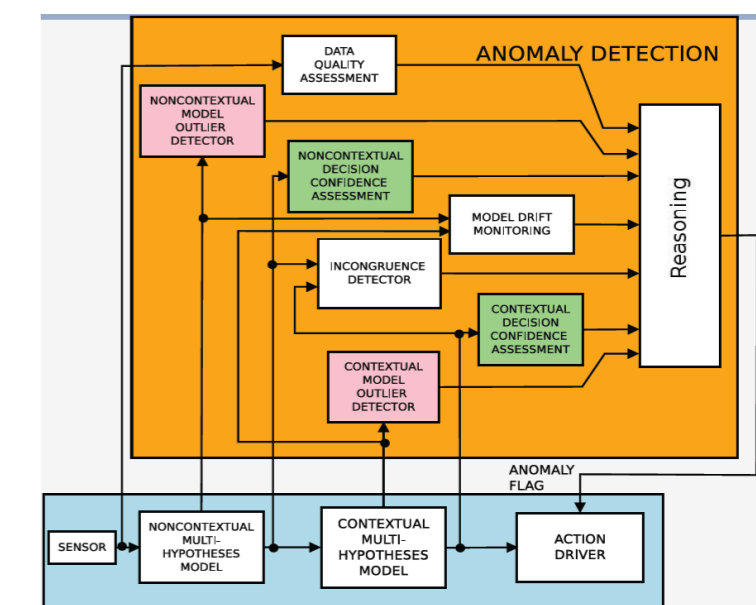


Statistical Anomaly Detection in Communication Networks

David J. Parish, Francisco J. Aparicio Navarro
 Loughborough University
 e-mail: {d.j.parish, elfja2}@lboro.ac.uk

Introduction

This Work Package proposes the design of an Anomaly Detection System with advance methodology for anomaly detection in battlespace scenarios. The aims of this work is to develop statistical algorithms for automatic detection and classification of anomalies from multidimensional, undersampled, non-complete datasets and unreliable networked battlespace environment sources, and to identify the nature and statistical characteristics of these anomalies.



Current Aims

- Current work focuses on the 'Data Quality Assessment' module, in the proposed Anomaly Detection System.
- Develop novel methodology that **automatically generates labelled network traffic datasets**, using the outcome of an unsupervised Intrusion Detection System (IDS).
- Apply **Feature Selection techniques** to the generated labelled dataset to automatically select the most appropriate set of metrics, and to automatically include new metrics.
- Incorporate information from different levels of the **domain knowledge** and **contextual information**.

Current IDS Capabilities

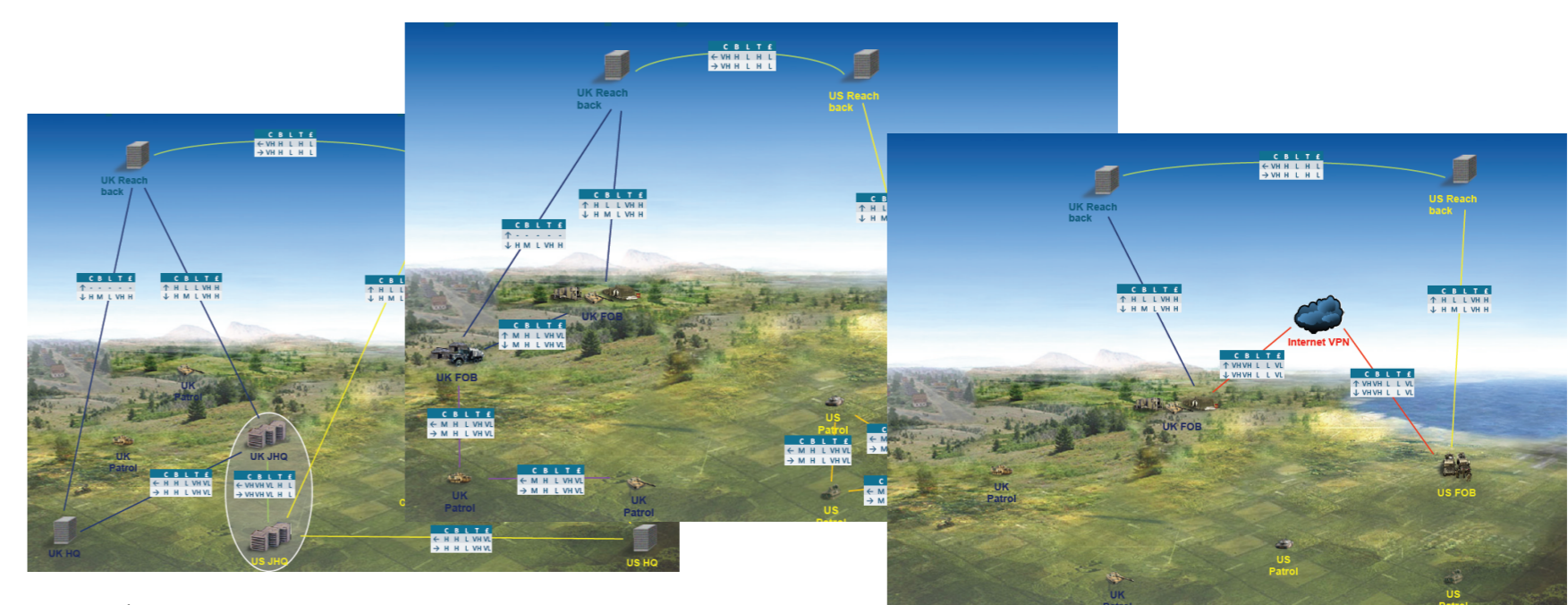
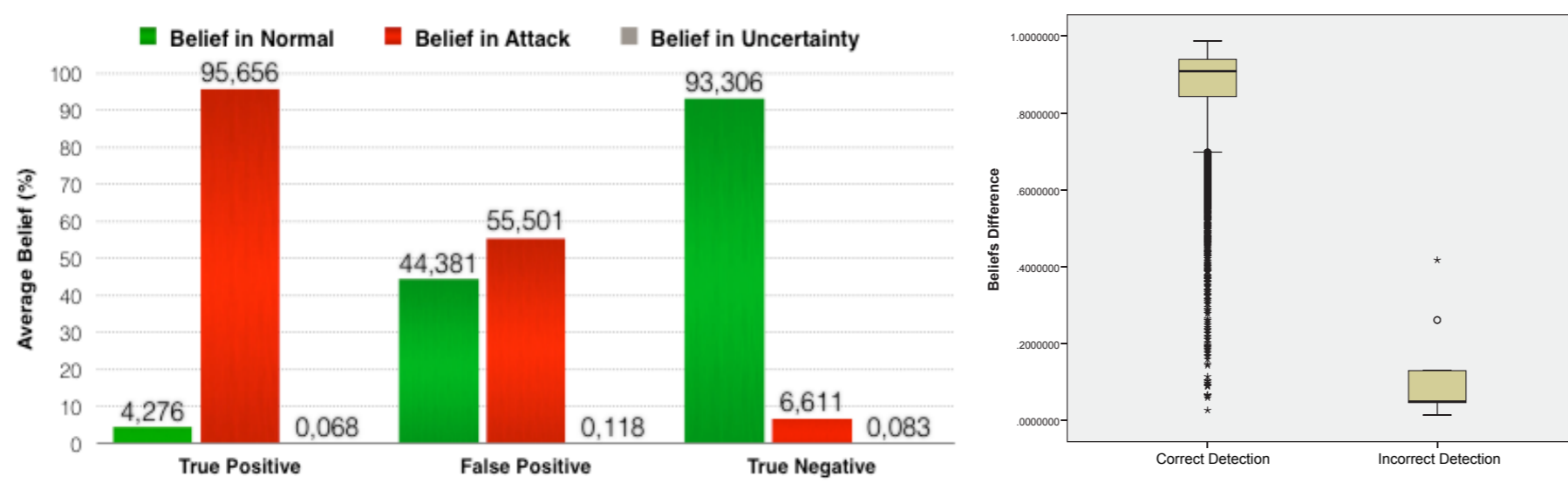
- **Protect** IEEE 802.11 networks from **Injection types of attacks**.
- Rogue frames cannot be detected by conventional approaches which look for known profiles in the WiFi frame content.
- A set of predefined metrics is isolated and a statistical reference is calculated to generate a baseline profile of normal traffic.
- For each frame and metric, the detection system provides levels of **belief in the three hypotheses Normal, Attack, Uncertainty**.
- Three innovative techniques are used to automatically generate the belief values, trained with 20-30 frames.
- The system uses a multi-layer approach. The beliefs are fused using **Dempster-Shafer Theory of Evidence**.

Automatic Data Labelling

- Frames in a dataset could be **labelled** according to the **final results of an unsupervised IDS**.
- Large difference between belief in *Normal* and *Attack* produces correct classifications – Strong Belief Results.
- Small difference between belief in *Normal* and in *Attack* produces misclassifications – Weak Belief Results.
- Considering **only strong belief results**, new datasets fully composed of correctly labelled instances are created.
- An efficient methodology to define the **boundary threshold** between strong and weak belief results has been found.

Contextual Information

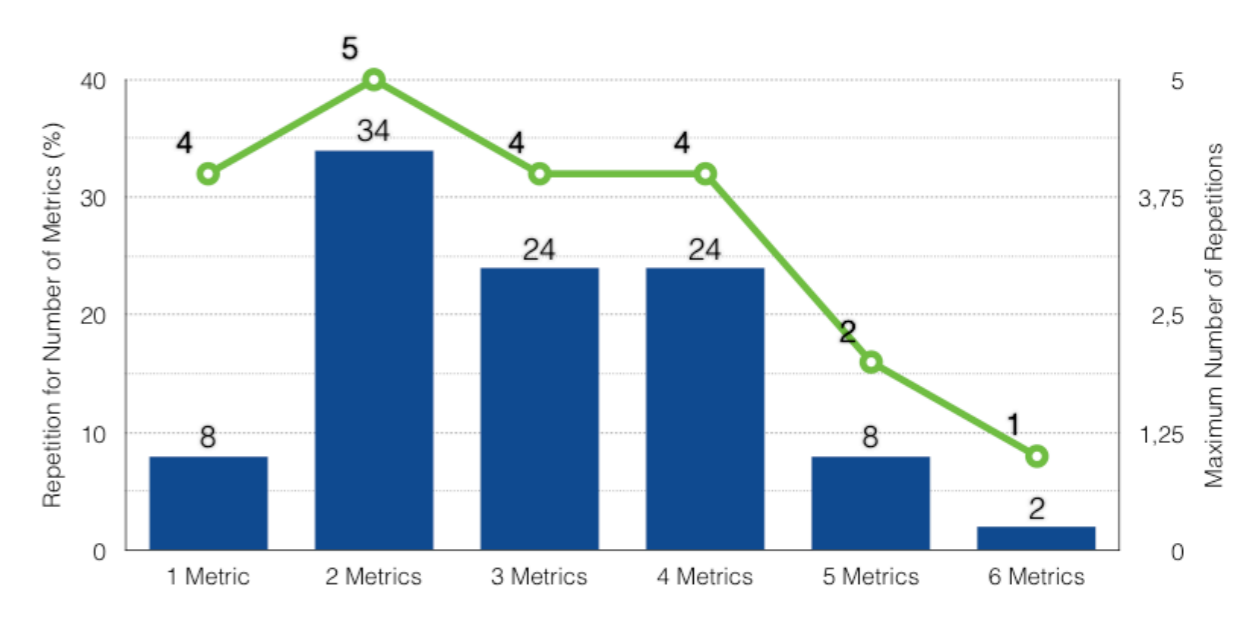
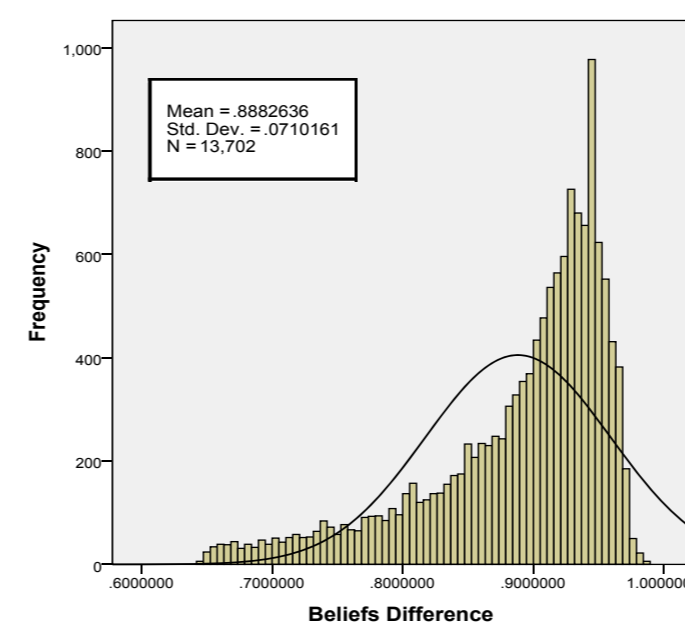
- The configuration of the Anomaly Detection System may be **modified according to** particular battlespace **environments**.
- Different missions generate **different network topologies**.
- Different topologies require/provide different parameters for the networked battlespace communications.
- These parameters could be used as a high-level source of **contextual information**.



* Dstl, 'Current and Future Network Topologies', 2013.

Feature Selection

- **Feature Selection** techniques automatically minimise and optimise the selection of metrics.
- A **Genetic Algorithm (GA)** based approach has been developed. GA replaces exhaustive approaches to optimise the selection of metrics.
- Processing the correctly labelled dataset with GA provides the set of metrics that produces the **best detection results**.



School of Electronic, Electrical and Systems Engineering

