Decision Fusion using Dempster-Schaffer Theory

Prof. D. J. Parish

High Speed networks Group
Department of Electronic and Electrical
Engineering
D.J.Parish@lboro.ac.uk

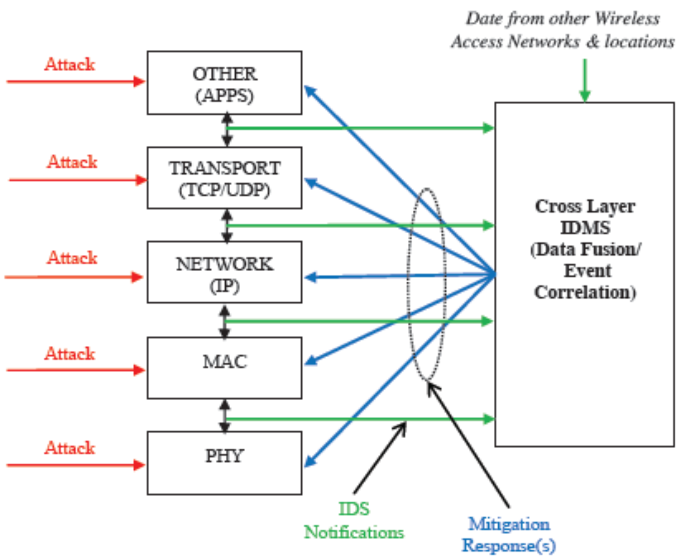
Loughborough University

Overview

- Introduction
- Why Fusion?
- Possible Approaches
  - Baysian
- Dempster-Schaffer Theory
  - Origin
  - Main Characteristics
- DS worked example
- DS Issues
  - Data Independence
  - Some issues with conflicting evidence
- Basic Belief Assignment
  - Possible Approaches
  - A “Light Weight” Approach
- Examples from Research

WHY DATA FUSION

- Multi-Metric or Cross-layer Anomaly Based IDSs outperform Single-metric detection results [3]
- Although there are cases in which IDSs that utilise information from a single metric might give good detection results, the presence of attacks is rarely accurately detectable by examining a single metric from one layer of the protocol stack.
- Multi-Metric IDSs combine information from two or more layers of the protocol stack
- The higher the number of metrics, the greater the chances to identify intrusions



DATA FUSION

- Data fusion:
  - Process of gathering information from multiple and heterogeneous sources and combining them towards obtaining a more accurate final result
  - The most common data fusion techniques
    - Bayesian Theory
    - Dempster-Shafer (D-S) Theory of Evidence
- **Bayesian** Theory
  - Calculates the probability of occurrence of a certain event, based on the experience extracted from previous events
  - Previous event probabilities is very difficult or impossible to determine
  - Does not directly assign probability to *uncertainty*

Dempster-Shafer Theory

- From Wikipedia, the free encyclopedia

- The **Dempster–Shafer theory** (**DST**) is a mathematical theory of evidence.[1] It allows one to combine evidence from different sources and arrive at a degree of belief (represented by a belief function) that takes into account all the available evidence. The theory was first developed by Arthur P. Dempster[2] and Glenn Shafer.[1][3]
- In a narrow sense, the term **Dempster–Shafer theory** refers to the original conception of the theory by Dempster and Shafer. However, it is more common to use the term in the wider sense of the same general approach, as adapted to specific kinds of situations. In particular, many authors have proposed different rules for combining evidence, often with a view to handling conflicts in evidence better

DATA FUSION

- **Dempster-Shafer** (D-S) Theory of Evidence
  - Mathematical discipline that combines evidences of information from multiple events to calculate the belief of occurrence of another event
  - PROS:
    - High potential for managing *Uncertainty*
    - Assigns probability to *Uncertainty*
    - Does not require a priori knowledge
    - Suitable for detecting previously unknown attacks
  - CONS:
    - Computation complexity increases exponentially with the number of possible event outcomes
    - Conflicting beliefs management assigning empty belief value
    - Evidences should be completely independent
- A comparative study of different data fusion methods is presented in [3]
- This work concludes that D-S theory is more promising than Bayesian in tasks of IDS

DEMPSTER-SHAFER

- Frame of Discernment Θ = {*Θ1, Θ2,..., Θn*}
- Finite set of all possible mutually exclusive outcomes about some problem domain
- All the observers must use the same frame of discernment

- 2^Θ, refers to every possible mutually exclusive subset of the elements of Θ
  - If Θ = {Attack, Normal}, then 2^Θ = {Attack, Normal, Uncertainty, Ø}

- Each subset is defined as an Hypothesis and receives a belief value within [0, 1]
- Assignment is known as the Basic Probability Assignment (BPA)

$$m : 2\Theta \rightarrow [0,1] \qquad \text{if} \begin{cases} m(\emptyset) = 0 \\ m(A) \geq 0, \forall A \subseteq \Theta \\ \sum_{A \subseteq \Theta} m(A) = 1 \end{cases}$$

- From the mass assignments, the upper and lower bounds of a probability interval can be defined. This interval contains the precise probability of a set of interest (in the classical sense), and is bounded by two non-additive continuous measures called **belief** (or **support**) and **plausibility**:
- The belief bel(*A*) for a set *A* is defined as the sum of all the masses of subsets of the set of interest:
- The plausibility pl(*A*) is the sum of all the masses of the sets *B* that intersect the set of interest *A*:
- The two measures are related to each other as follows:
- And conversely, for finite *A*, given the belief measure bel(*B*) for all subsets *B* of *A*, we can find the masses m(*A*) with the following inverse function:
- where $|A - B|$ is the difference of the cardinalities of the two sets.[4]

DEMPSTER-SHAFER - EXAMPLE

Sensor 1	
Attack	0.32
Normal	0.25
Uncertainty	0.43

Sensor 2	
Attack	0.35
Normal	0.1
Uncertainty	0.55

$$m(E) = \frac{\sum_{X \cap Y = E} m_1(X) * m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X) * m_2(Y)} \quad \forall E \neq \emptyset$$

$\mathbf{m_2 \backslash m_1}$	$\{A\}$: 0.32	$\{N\}$: 0.25	$\{A, N\}$: 0.43
$\{A\}$: 0.35	0.11	0.09	0.15
$\{N\}$: 0.1	0.03	0.025	0.04
$\{A, N\}$: 0.55	0.18	0.14	0.24

$$m(A) = 1.136 * (0.11 + 0.15 + 0.18) = 0.5$$
$$m(N) = 1.136 * (0.025 + 0.04 + 0.14) = 0.233$$
$$m(A|N) = 1.136 * (0.24) = 0.272$$

DEMPSTER-SHAFER - EXAMPLE

Sensor 1	
Attack	0.32
Normal	0.25
Uncertainty	0.43

Sensor 2	
Attack	0.35
Normal	0.1
Uncertainty	0.55

$$m(E)=\frac{\sum_{X\cap Y=E} m_1(X)*m_2(Y)}{1-\sum_{X\cap Y=\emptyset} m_1(X)*m_2(Y)}\quad \forall E\neq\emptyset$$

$m_2\backslash m_1$	$\{A\}$: 0.32	$\{N\}$: 0.25	$\{A,N\}$: 0.43
$\{A\}$: 0.35	0.11	0.09	0.15
$\{N\}$: 0.1	0.03	0.025	0.04
$\{A,N\}$: 0.55	0.18	0.14	0.24

$$m(A) = 1.136 * (0.11 + 0.15 + 0.18) = 0.5$$
$$m(N) = 1.136 * (0.025 + 0.04 + 0.14) = 0.233$$
$$m(A|N) = 1.136 * (0.24) = 0.272$$

DEMPSTER-SHAFER - EXAMPLE

Sensor 1	
Attack	0.32
Normal	0.25
Uncertainty	0.43

Sensor 2	
Attack	0.35
Normal	0.1
Uncertainty	0.55

$$m(E)=\frac{\sum_{X\cap Y=E} m_1(X)*m_2(Y)}{1-\sum_{X\cap Y=\emptyset} m_1(X)*m_2(Y)} \quad \forall E\neq\emptyset$$

$m_2 \backslash m_1$	$\{A\}$: 0.32	$\{N\}$: 0.25	$\{A,N\}$: 0.43
$\{A\}$: 0.35	0.11	0.09	0.15
$\{N\}$: 0.1	0.03	0.025	0.04
$\{A,N\}$: 0.55	0.18	0.14	0.24

$$m(A) = 1.136 * (0.11 + 0.15 + 0.18) = 0.5$$
$$m(N) = 1.136 * (0.025 + 0.04 + 0.14) = 0.233$$
$$m(A|N) = 1.136 * (0.24) = 0.272$$

DEMPSTER-SHAFER - EXAMPLE

Sensor 1	
Attack	0.32
Normal	0.25
Uncertainty	0.43

Sensor 2	
Attack	0.35
Normal	0.1
Uncertainty	0.55

$$m(E)=\frac{\sum_{X\cap Y=E} m_1(X)*m_2(Y)}{1-\sum_{X\cap Y=\emptyset} m_1(X)*m_2(Y)}\quad \forall E\neq\emptyset$$

$m_2 \backslash m_1$	$\{A\}: 0.32$	$\{N\}: 0.25$	$\{A,N\}: 0.43$
$\{A\}: 0.35$	0.11	0.09	0.15
$\{N\}: 0.1$	0.03	0.025	0.04
$\{A,N\}: 0.55$	0.18	0.14	0.24

$$m(A) = 1.136 * (0.11 + 0.15 + 0.18) = 0.5$$
$$m(N) = 1.136 * (0.025 + 0.04 + 0.14) = 0.233$$
$$m(A|N) = 1.136 * (0.24) = 0.272$$

An example with more sensors

		Sensor					
		#1	*#2*	*#3*	*#4*	*#5*	*#6*
Hypothesis	*Normal*	0.3	0.217	0.667	0.667	0.217	0.217
	Attack	0.4	0.567	0.167	0.167	0.567	0.567
	Uncertainty	0.3	0.216	0.166	0.166	0.216	0.216

		Iteration				
		#1 - #2	*R - #3*	*R - #4*	*R - #4*	*Final Results*
Hypothesis	*Normal*	0.262	0.857	0.187	0.746	0.475
	Attack	0.65	0.107	0.751	0.247	0.524
	Uncertainty	0.088	0.036	0.062	0.007	0.001

BASIC PROBABILITY ASSIGNMENT

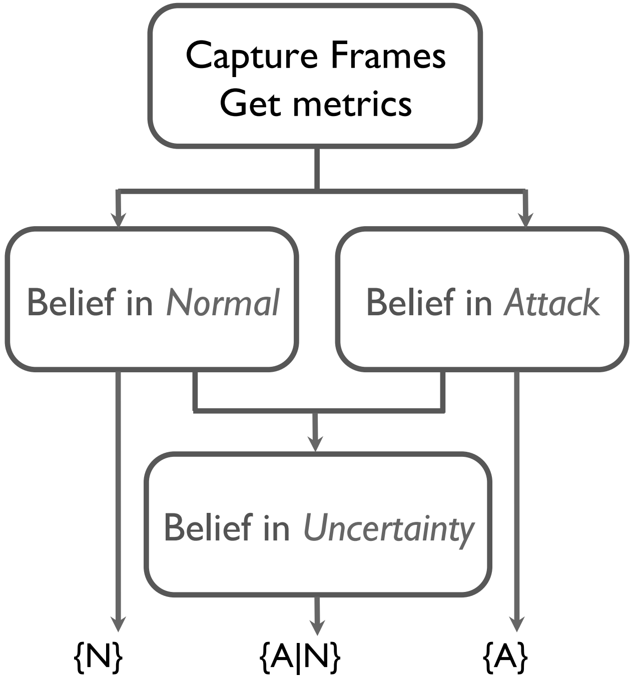
- **Current Techniques**
  - Empirical approach
  - Expert opinion

  - Manually assignment
  - Fixed Thresholds
  - Fixed Scales
  - Fixed Linear functions
    - Unable to automatically adapt without IDS administrator

  - Data Mining techniques
    - Require Gathering data, Processing, Training, Perform analysis, etc.
      - Unable to automatically adapt in Real-Time

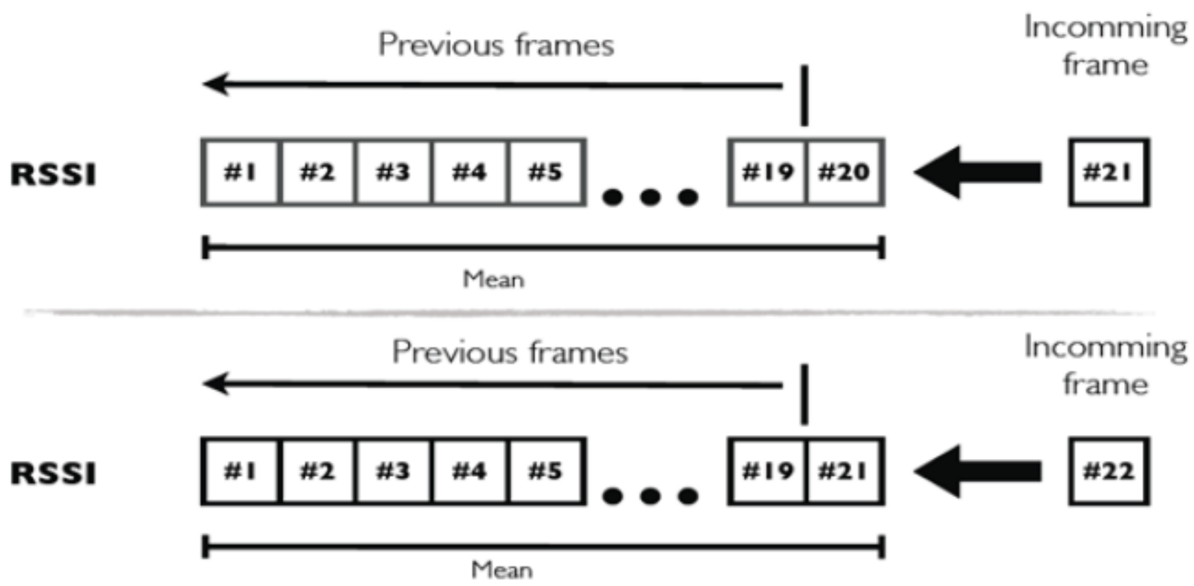
BASIC PROBABILITY ASSIGNMENT - METHODOLOGY

- We proposed a novel BPA methodology [4]
  - Three independent Statistical approaches
  - Automatically adapt detection capabilities
  - No intervention from IDS administrator
  - Light weight profiling process
  - Tested with diverse number of Wireless Network Attacks



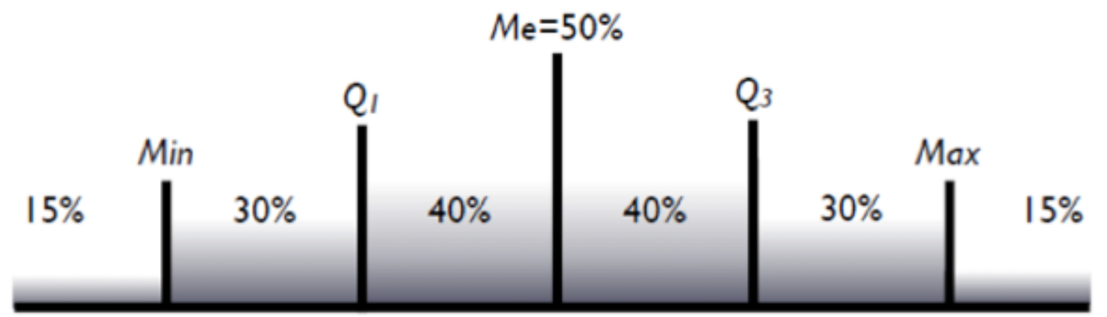
BASIC PROBABILITY ASSIGNMENT - METHODOLOGY

- Sliding window of ~30 frames
- If current frame is Legal → Slides
- If current frame is Malicious → Drops the frame



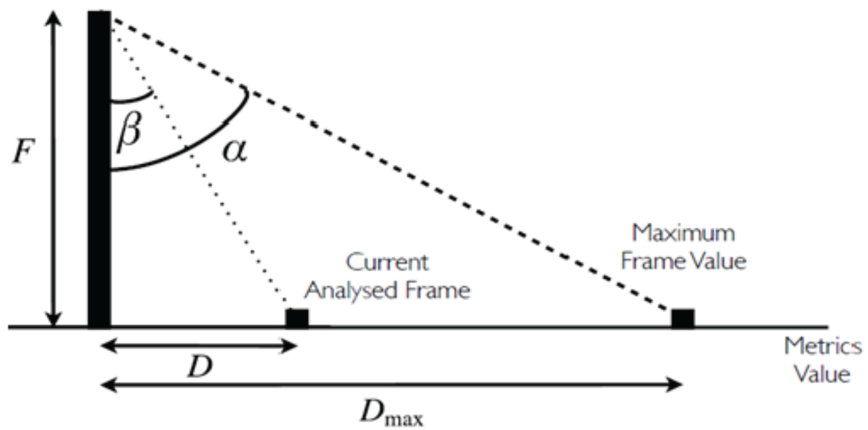
BASIC PROBABILITY ASSIGNMENT - BELIEF IN NORMAL

- Degree of dispersion of Dataset
- Similar to Boxplot method
- Quartiles define the scales boundaries
- Length of scales varies
    - Automatically adjust to the network behaviour changes



BASIC PROBABILITY ASSIGNMENT - BELIEF IN ATTACK - ANGLE

- Frequency and Euclidean Distance
- Mean or Mode - Reference point
- Angle α - Reference of maximum belief
- Angle β - Reference of belief to current analysed frame
- Lineal function between α and β generates belief in Attack



BASIC PROBABILITY ASSIGNMENT - BELIEF IN UNCERTAINTY

- Belief in *Uncertainty* is used as adjustment value
- Provisional *Uncertainty* value:

$$Belief_{(Unc.)} = \frac{0.5 \cdot Belief_{Min}}{Belief_{Max}}$$

- Condition of D-S Theory:

$$\sum_{A \subseteq \Theta} m(A) = 1$$

- Adjustment value:

$$\mu = \frac{(X - 1)}{3}$$

- *X* = Summation of the three beliefs

Potential Problems

- All data and sensors used by DS should be independent.
- This is difficult to achieve in practice and there is considerable literature to indicate that total independence is not always required in practice.
- Misleading results can be generated if there is contradictory evidence, or certain values are 0.

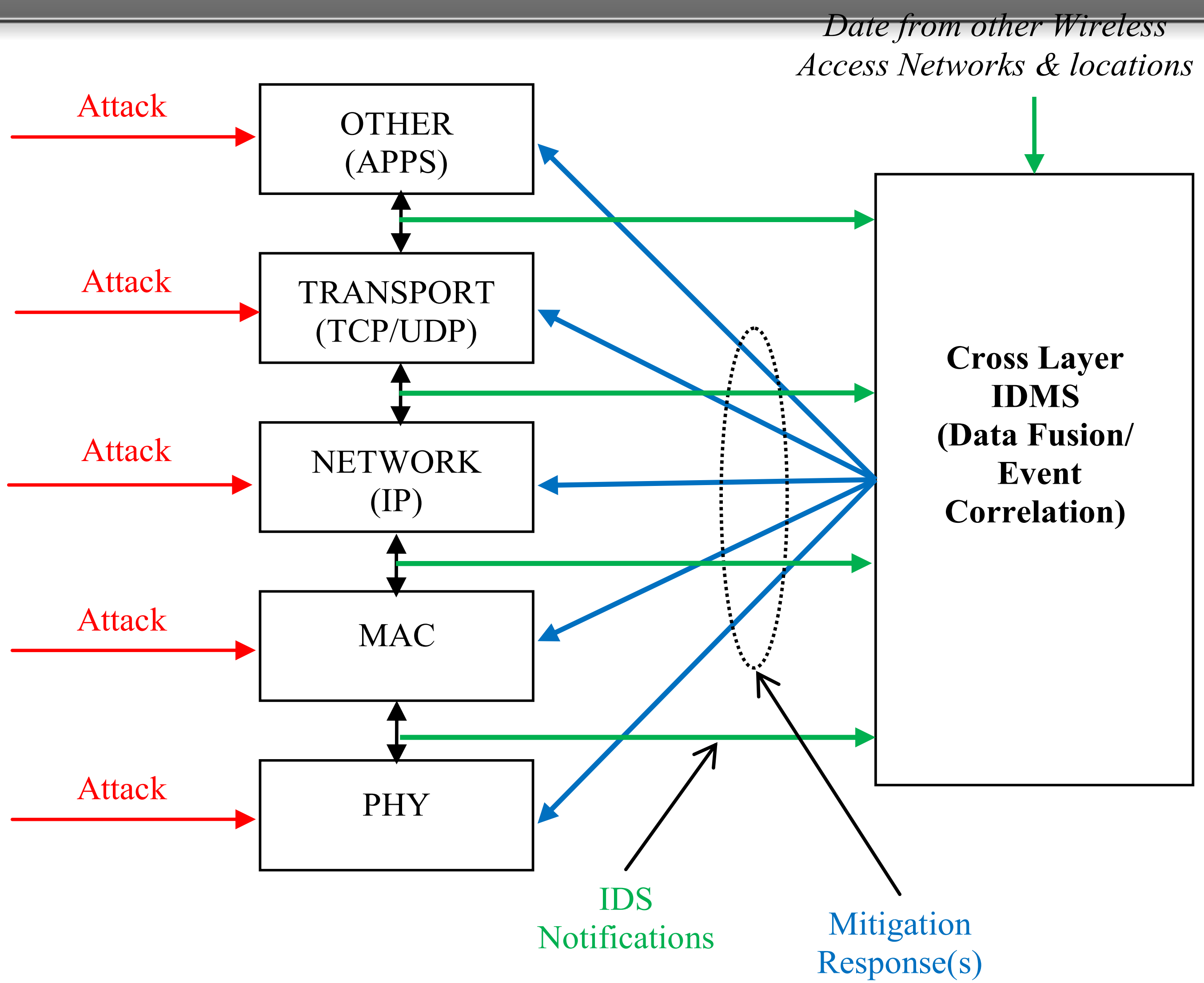
% = {A,B,C}
m1 = {A} (0.99), {B} (0.01); [{C} (0)]
m2 = {C} (0.99), {B} (0.01); [{A} (0)]

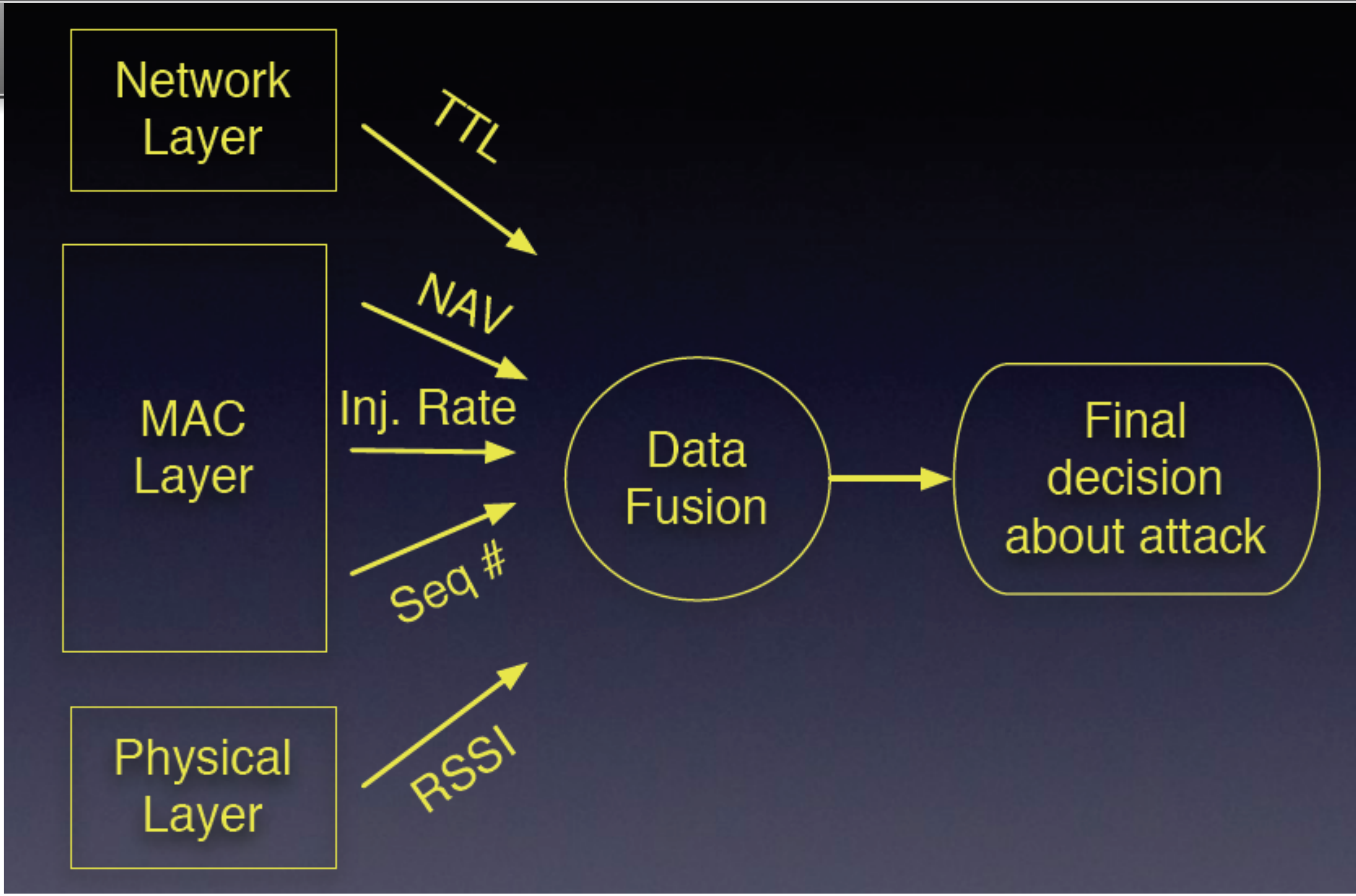
m1 + m2 = {B} (1.0)

In some cases this may be sensible, in others it will not be.

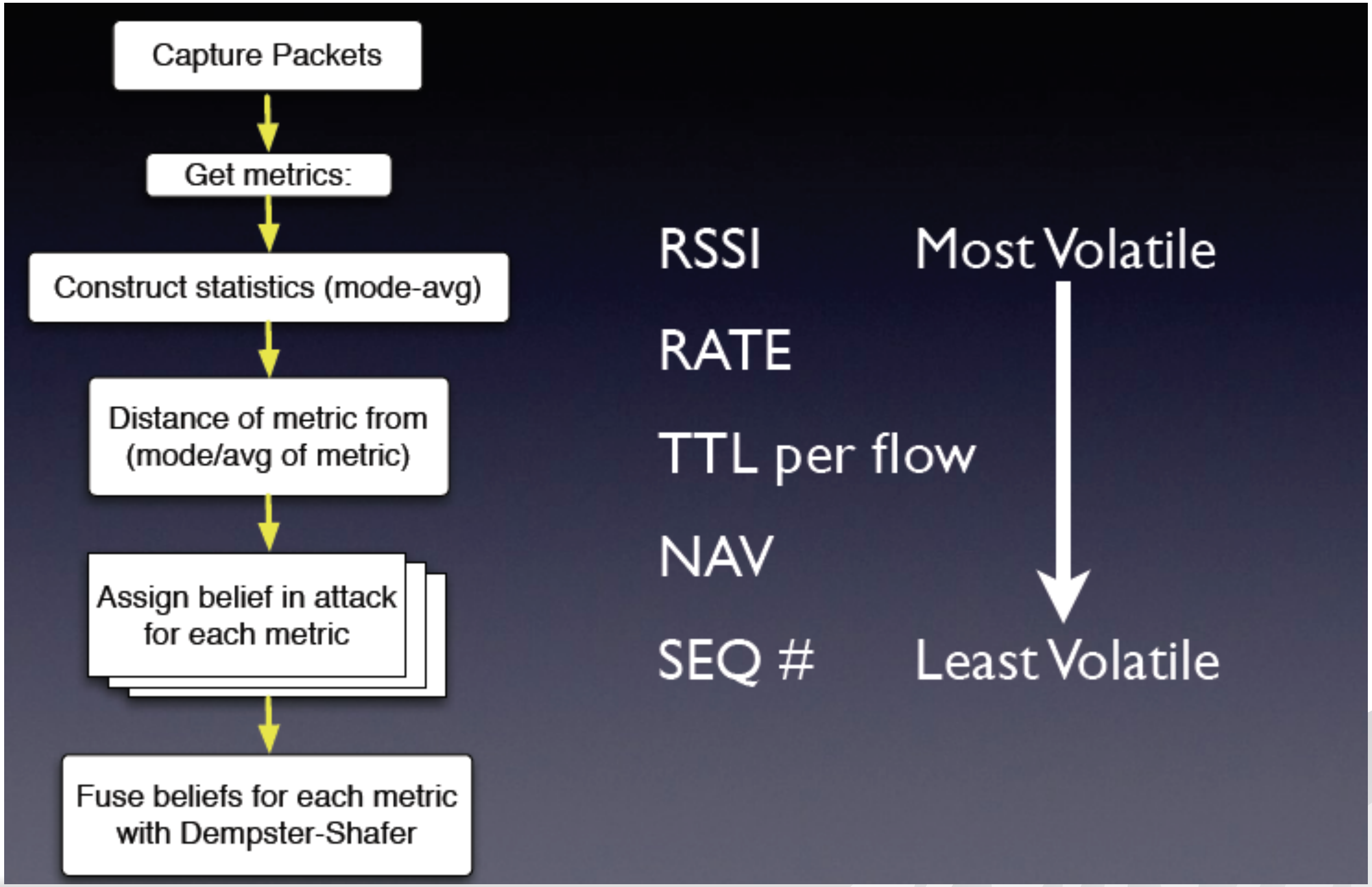
Mixed Layer Abuse Detection in Wireless Networks

- Aims to use multiple metrics from different layers to improve abuse detection in Wireless networks.
- Data Fusion based on Dempster-Schaffer theory of evidence.
    - Data Mining approaches could be evaluated


Date from other Wireless
Access Networks & locations
Attack
OTHER
(APPS)
Attack
TRANSPORT
(TCP/UDP)
Attack
NETWORK
(IP)
Attack
MAC
Attack
PHY
Cross Layer
IDMS
(Data Fusion/
Event
Correlation)
IDS
Notifications
Mitigation
Response(s)


Network Layer
TTL
NAV
MAC Layer
Inj. Rate
Seq #
Physical Layer
RSSI
Data Fusion
Final decision about attack

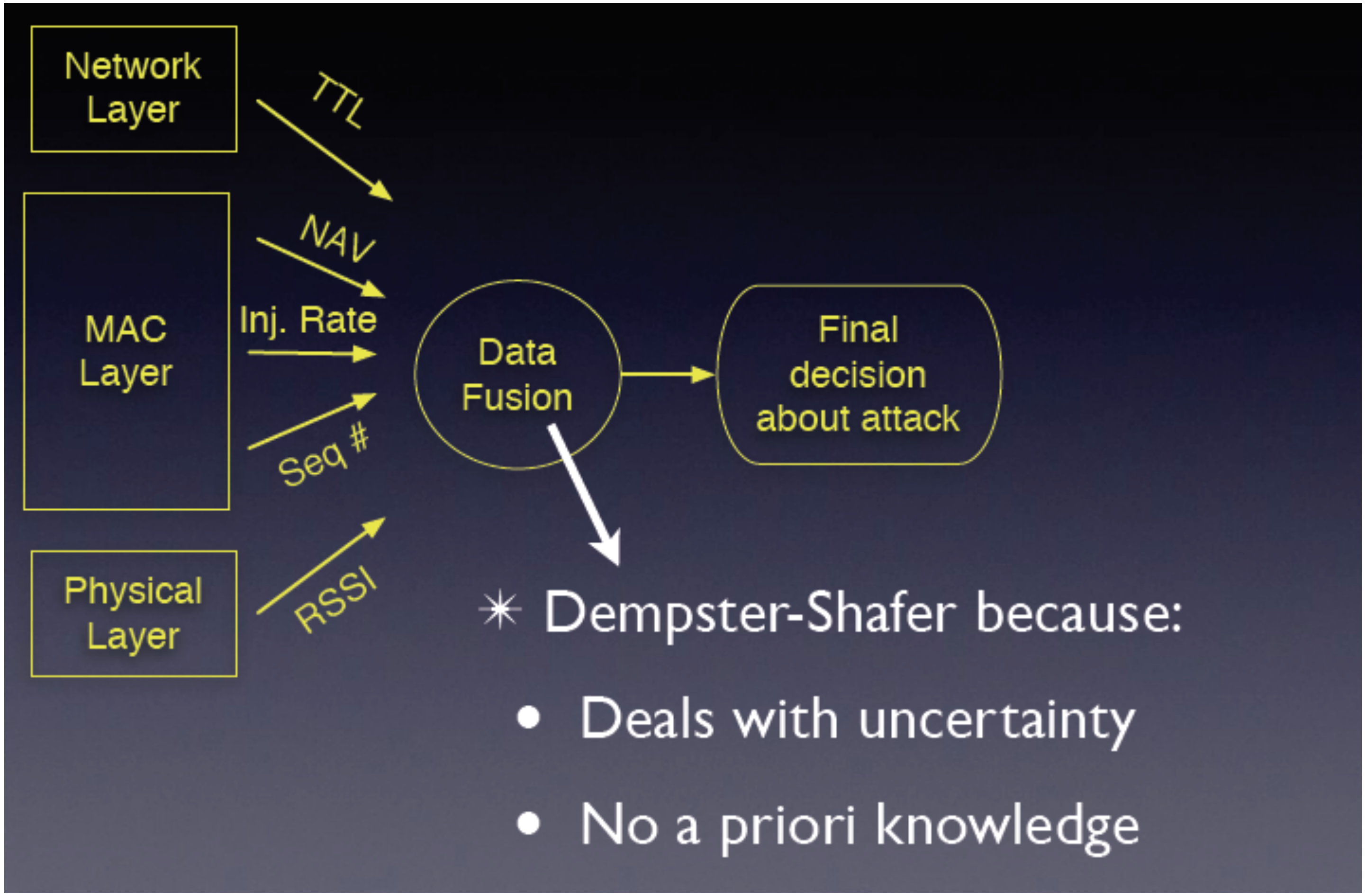
Methodology



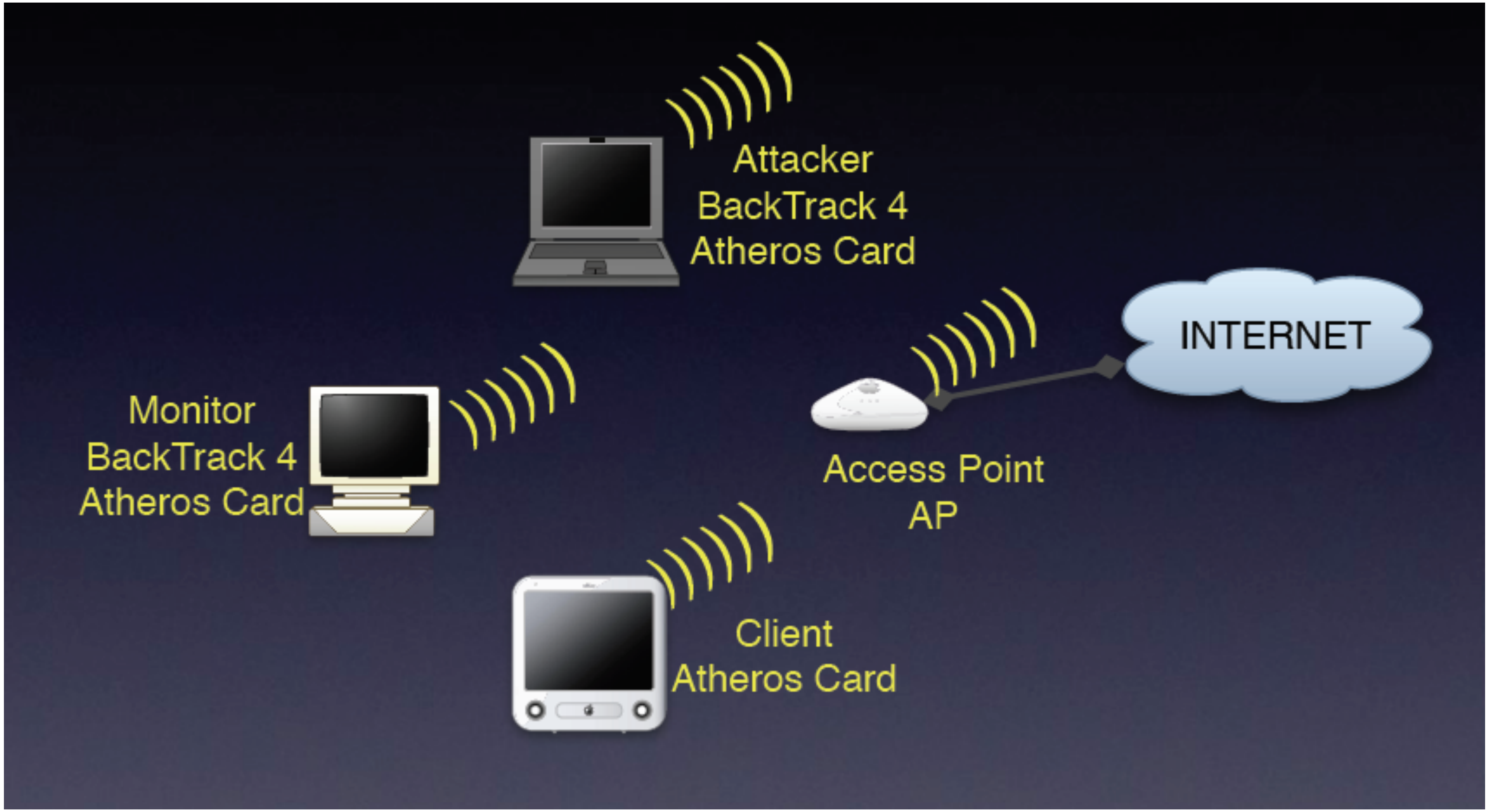
EPSRC
Pioneering research and skills

Loughborough University

Data Fusion



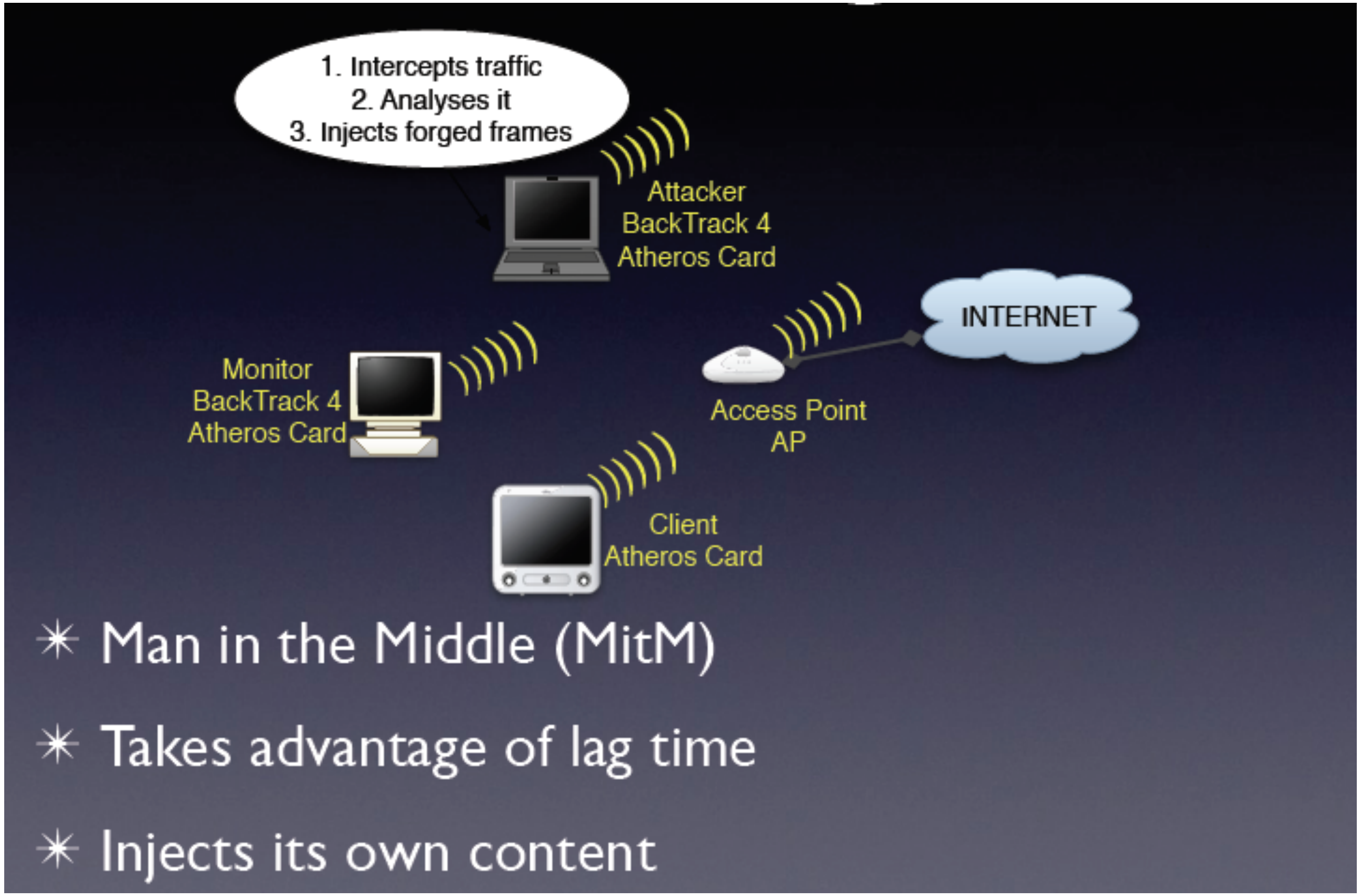
Testbed


Attacker
BackTrack 4
Atheros Card
Monitor
BackTrack 4
Atheros Card
INTERNET
Access Point
AP
Client
Atheros Card

EPSRC Pioneering research and skills

Loughborough University

Man In The Middle Attack



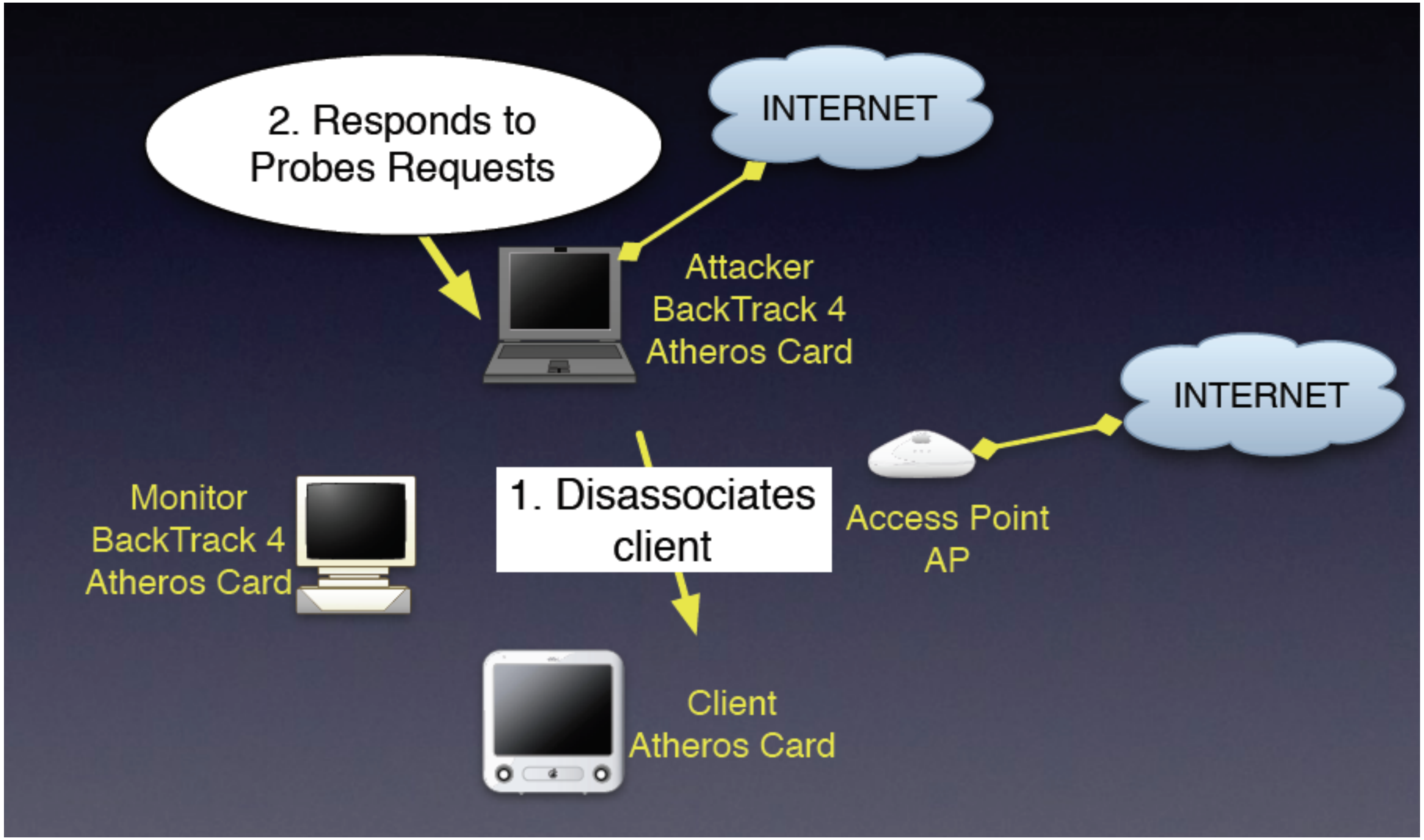
EPSRC
Pioneering research and skills

Loughborough University

Man-In-The Middle Attack Results

Metrics	Type	%	Result %
NAV + SEQ	FN	0	0
	FP	7/63	11.1
RSSI + NAV + SEQ	FN	0	0
	FP	8/63	12.7
RSSI + TTL + RATE	FN	0	0
	FP	0	0
All metrics	FN	0	0
	FP	0	0

Rogue Access Point



Rogue Access Point Attacks

Method	Rate	ESSID Spoof
Airbase	Fixed at 1Mbps	No
Airbase -a	Fixed at 1Mbps	Yes
Host AP	Normal Rate	No

Rogue Access Point Results

Metrics	Type	Airbase	Airbase ESSID Spoof	HostAP
NAV + SEQ	Detected ?	Yes	Yes	Yes
	FP	0/405	0/246	0/57
RSSI + NAV + SEQ	Detected ?	Yes	Yes	Yes
	FP	35/405	2/246	3/57
RSSI + TTL + RATE	Detected ?	No	Yes	No
	FP	100%	0/246	100%
All metrics	Detected ?	Yes	Yes	Yes
	FP	0/405	0/246	0/57

EPSRC
Pioneering research and skills

Loughborough University

Benefits of Extra Metrics

No. of Metrics	Beliefs		
	Attack	No Attack	Uncertainty
NAV-SEQ	0.569	0.314	0.118
RSSI - NAV - SEQ	0.664	0.263	0.073
RSSI - TTL - Rate	0.575	0.329	0.096
5 metrics	0.710	0.272	0.018

Summary and Conclusions