University of Glasgow

The Past is no Predictor of the Future

Black Swans, Artificial Intelligence, Cyber Security and the End of Risk Assessment in Air Traffic Management

Belgocontrol, Friday 29th Septermber 2017,

Prof. Chris Johnson,

School of Computing Science, University of Glasgow, Scotland.

http://www.dcs.gla.ac.uk/~johnson

University of Glasgow

Overview

- New uncertainties in Air Traffic Management.
- “Black swan” events seem more common.
- Artificial Intelligence creates new possibilities.
- Cyber security is an increasing concern.

Overview: take Home Message

- Three challenges:
  - Black Swans, Artificial Intelligence, Cyber Security.

Overview: take Home Message

- Three challenges:
  - Black Swans, Artificial Intelligence, Cyber Security.
- One common concern:
  - The Death of Risk Assessment.

Overview: take Home Message

- Three challenges:
  - Black Swans, Artificial Intelligence, Cyber Security.
- One common concern:
  - The Death of Risk Assessment.
- One focus for technical innovation:
  - How do we sustain hazard analysis?
  - Can we engineer what “we know we don’t know”.

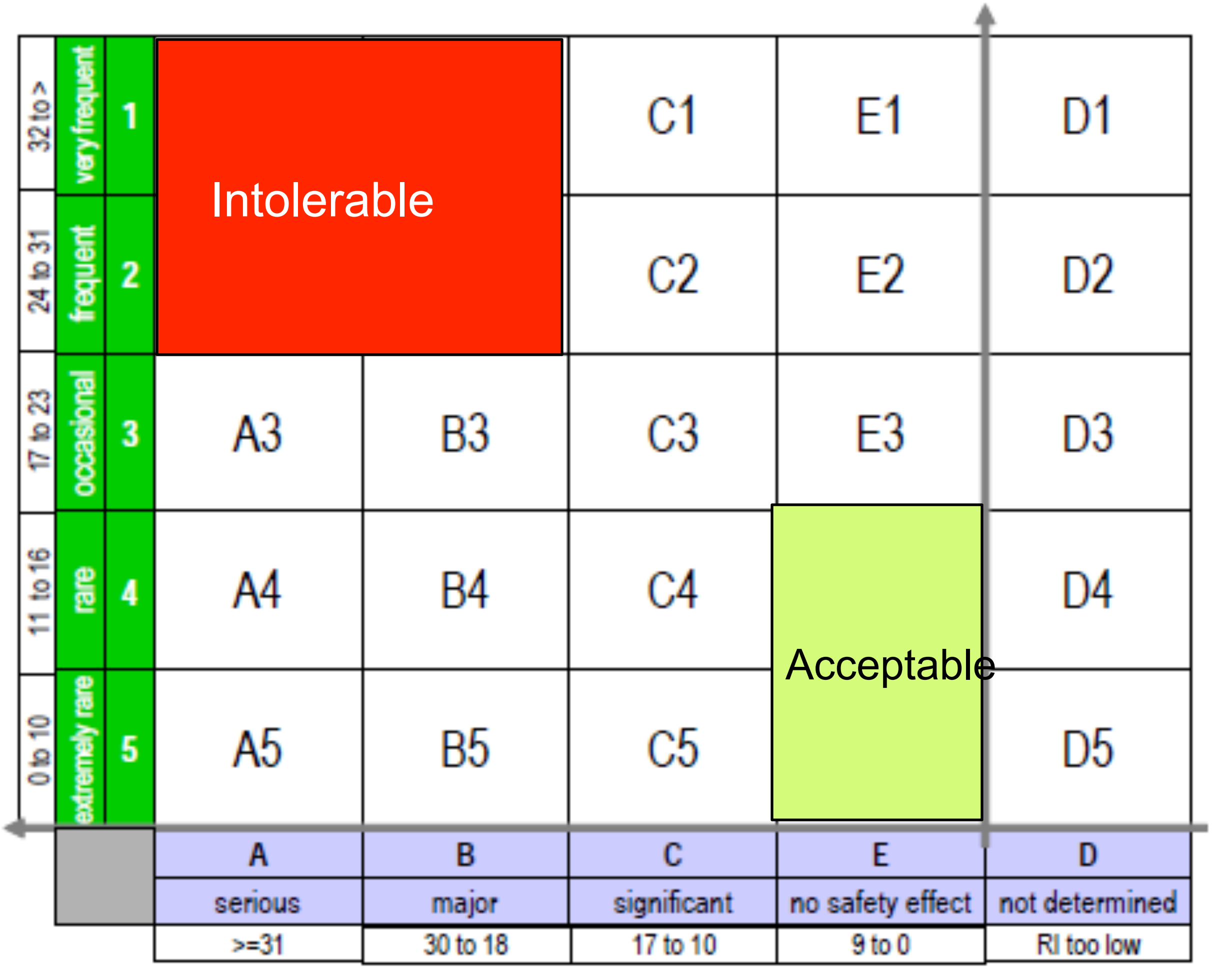
Or put another way

- In Air Traffic Management
- Past No Longer Valid for Predicting Future…
- So what can we do?

Risk Assessment

- SES CR 2096/2005 (1035/2011) ANSPs must reduce risk ‘as far as reasonably practicable’
- ‘risk’ means the combination of the overall probability, or frequency of occurrence of a harmful effect induced by a hazard and the severity of that effect; (CE IR 1035/2011)
- ‘hazard’ means any condition, event, or circumstance which could induce an accident;

Risk Matrices



32 to >	very frequent	1	Intolerable		C1	E1	D1
24 to 31	frequent	2			C2	E2	D2
17 to 23	occasional	3	A3	B3	C3	E3	D3
11 to 16	rare	4	A4	B4	C4	Acceptable	D4
0 to 10	extremely rare	5	A5	B5	C5		D5
			A	B	C	E	D
			serious	major	significant	no safety effect	not determined
			>=31	30 to 18	17 to 10	9 to 0	RI too low

Risk

$$Risk = \sum_{h=1}^{n} (probability_h \, \mathrm{x} \, consequence_h)$$

- Depends on hazard analysis.
- Structured common sense:
  - FMECA – failure modes;
  - HAZOPs – guide words.

Existing Challenges

- Risk assessment fails for software:
  - Cannot estimate probability of bugs;
  - IEC61508, ED-153 rely on ‘tricks’;
  - Very few people understand SILs, SWALs etc.
- Risk assessment fails for human factors:
  - Very few are happy with HRA;
  - Some claim it is “psychologically vacuous”;
  - Largely determined by context (PSFs).
- Almost impossible to validate.

Challenges for Risk Assessment

	Governmental	Organisational	Individual	Technical
Black Swans	What does 'acceptably safe' mean for Black Swan events?	How to manage finite resources to plan for very rare events?	How to mitigate human contribution to risks we never experienced?	How to ensure sufficient range of 'black swan' scenarios are considered?
Artificial intelligence	How to promote industry and innovation without exposing society to risk?	How to show systems that emulate human cognitive behavior acceptably safe?	How to help operators interact with autonomous systems?	How to test non-deterministic autonomous systems?
Cyber security	How to protect public and dissuade other nations from attacking?	How much to invest when the risk changes and is uncertain?	How to assess the human contribution to security?	How to protect systems when the past is no predictor of future risks?

Black Swans and Philosophy of Induction

- Hume's uniformity of nature;
  - Don't know chemical reason why emeralds are green;
  - Cause based on induction not reason/deduction.

- Leads to fundamental problem:
  - Assume you will only see white swans
  - Shows limits of learning from induction.

Taleb's Black Swan Events

- Nassim Nicholas Taleb:
  - Statistician, journalist, author, academic;
  - Critic of conventional risk management.
- A black swan event:
  - deviates beyond normal expectation in situation;
  - hence is extremely difficult to predict;
  - tend to have a disproportionate impact.
- Make society robust against BS events:
  - "Convex tinkering" decentralized enquiry;
  - Better than directed research programmes.

Rumsfeld Explains Black Swans

Imperfect Information for ATSEPs

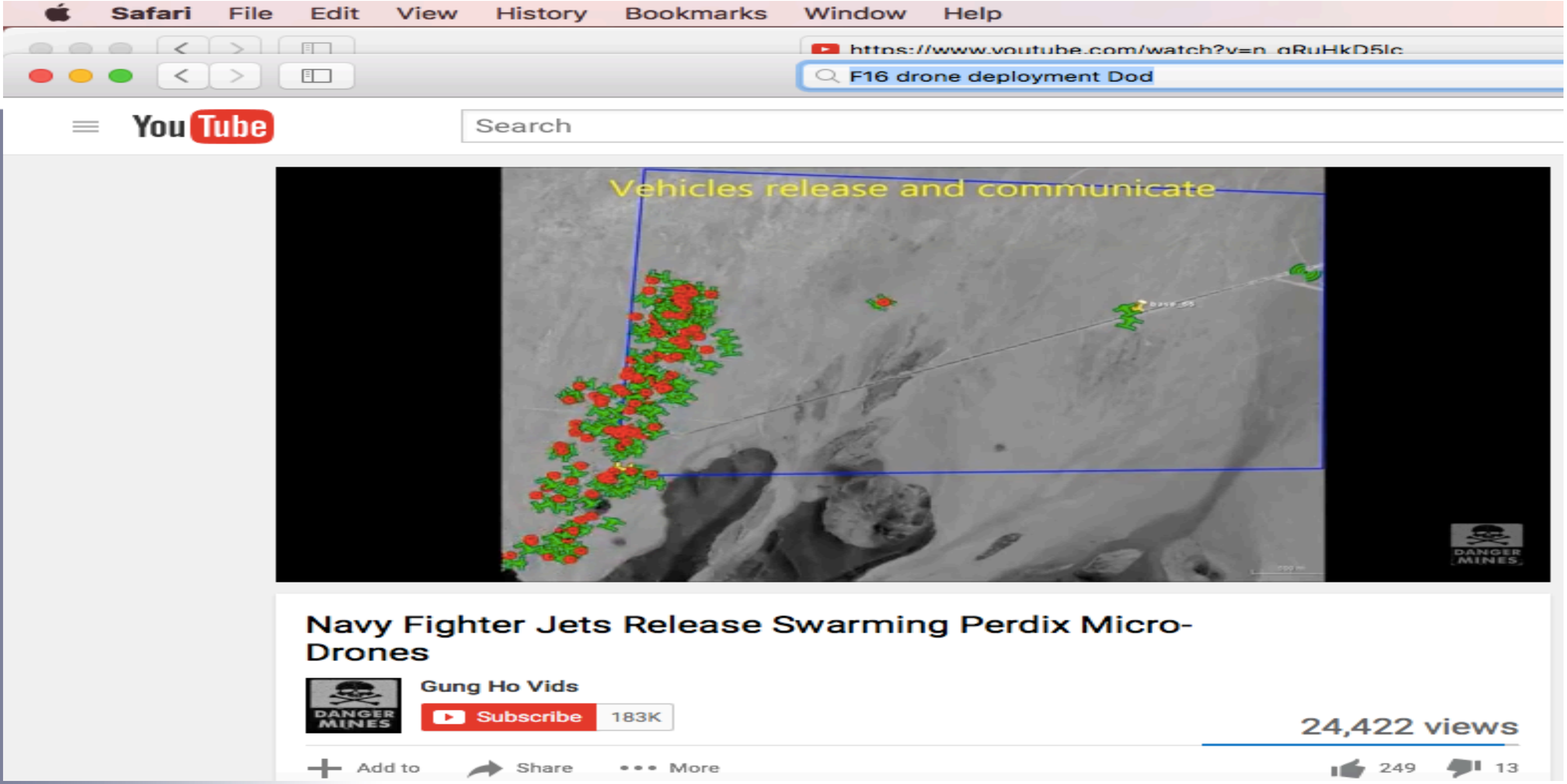
- Accumulator battery based UPS:
  - few seconds before generator starts;
  - Lightning cause surges across national grid.
- Power keeps tripping, blows UPS protection;
  - Batteries keep being used with each surge;
  - Batteries not recharging between surges;
  - ANSP can gradually see UPS failing.
- Eventually, power trips with no UPS backup.

What Black Swans Affect SESAR Deployment?

Challenges for Risk Assessment

	Organisational	Individual	Technical
Black Swans	What does 'acceptably safe' mean for Black Swan events?	How to mitigate human contribution to risks we never experienced?	How to ensure sufficient range of 'black swan' scenarios are considered?
Artificial intelligence	How to show systems that emulate human cognitive behavior acceptably safe?	How to help operators interact with autonomous systems?	How to test non-deterministic autonomous systems?
Cyber security	How much to invest when the risk changes and is uncertain?	How to assess the human contribution to security?	How to protect systems when the past is no predictor of future risks?

Perdix



https://www.youtube.com/watch?v=n_qRuHkD5Ic

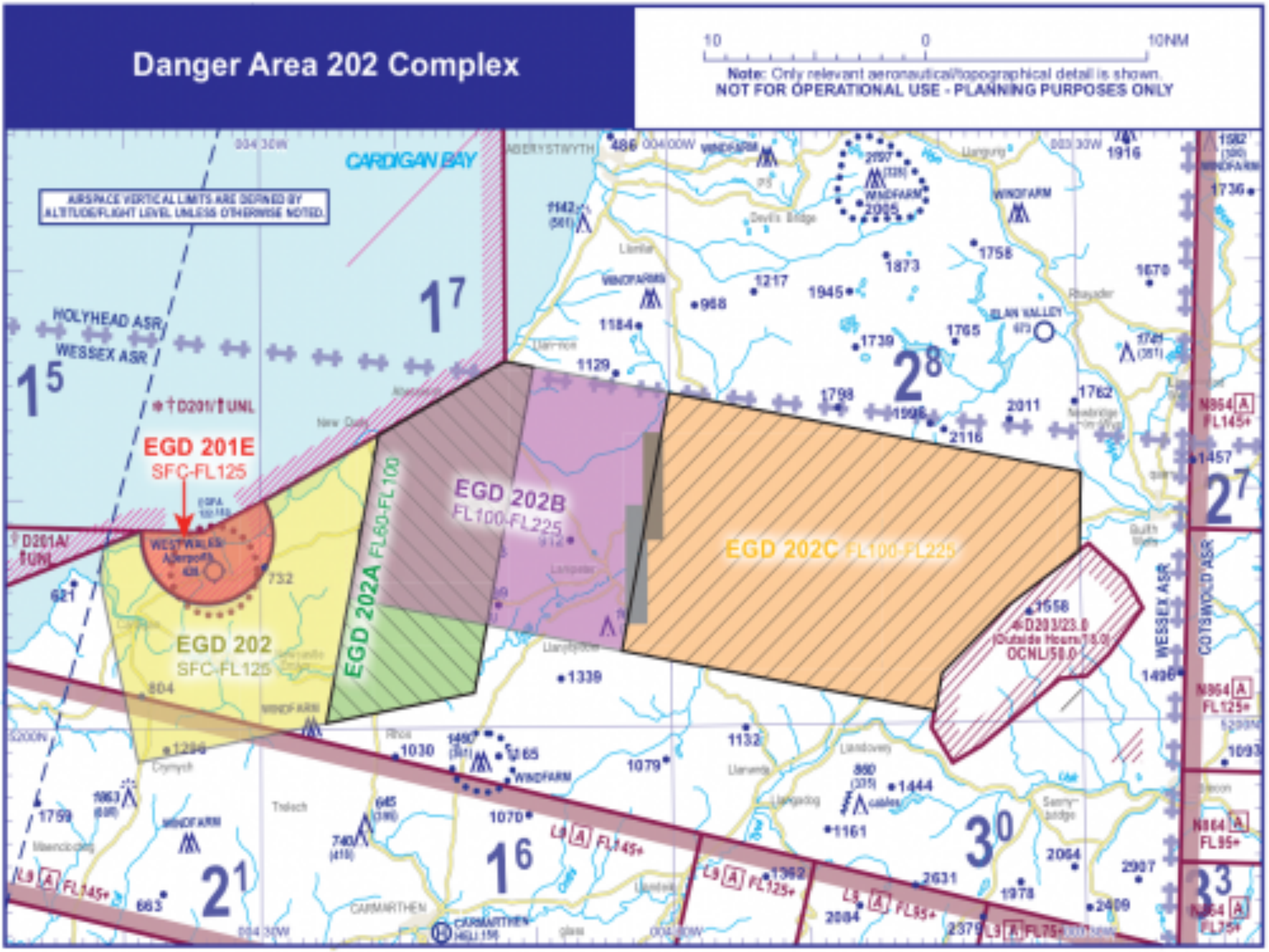
F16 drone deployment Dod

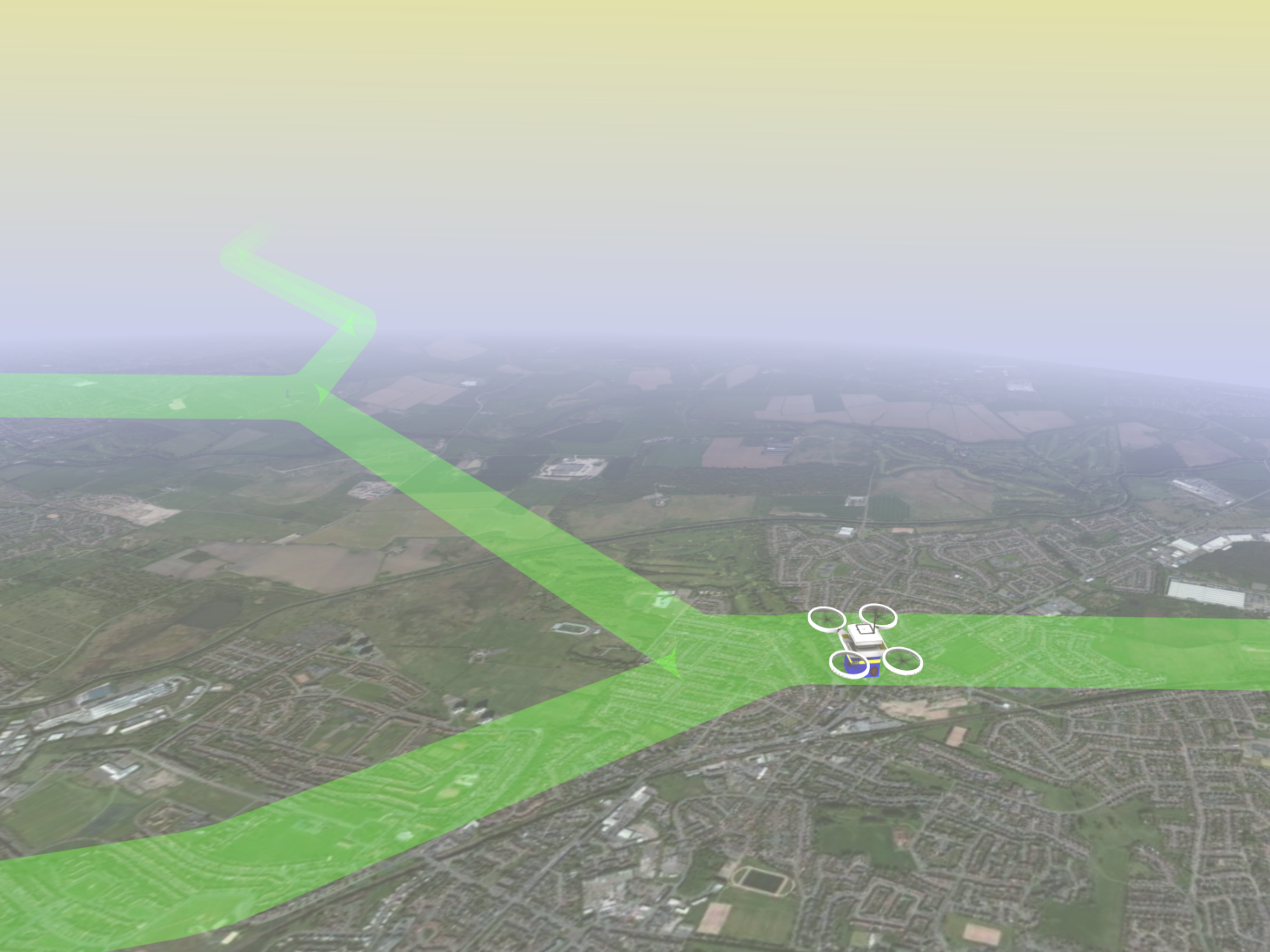
Navy Fighter Jets Release Swarming Perdix Micro-Drones

Gung Ho Vids

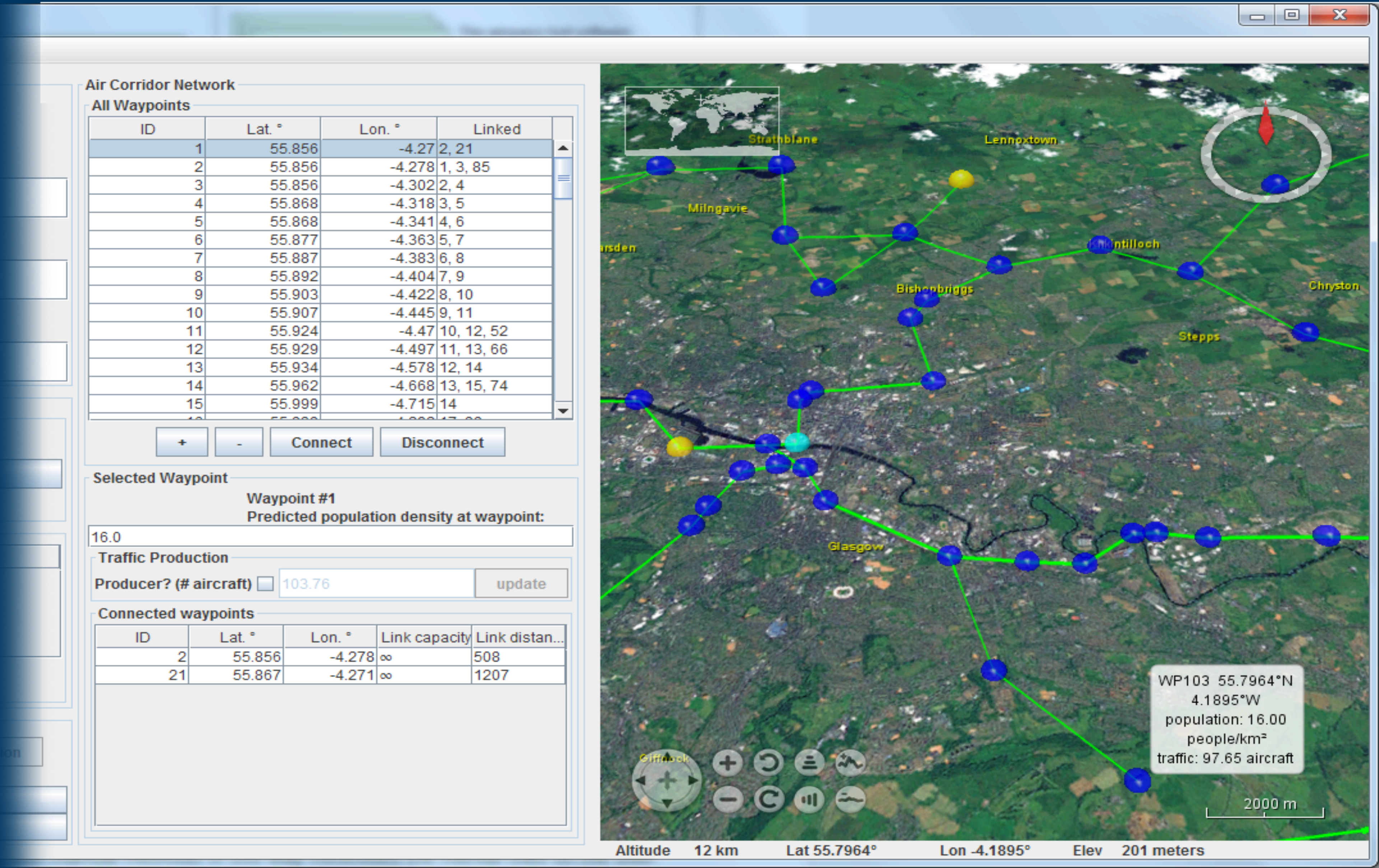
24,422 views

https://www.youtube.com/watch?v=0WNNanoUu2I


Danger Area 202 Complex
10 0 10NM
Note: Only relevant aeronautical/topographical detail is shown.
NOT FOR OPERATIONAL USE - PLANNING PURPOSES ONLY
CARDIGAN BAY
AIRSPACE VERTICAL LIMITS ARE DEFINED BY ALTITUDE/FLIGHT LEVEL UNLESS OTHERWISE NOTED.
HOLYHEAD ASR
WESSEX ASR
D201/UNL
EGD 201E
SFC-FL125
WEST WALES Airport
D201A/ UNL
EGD 202
SFC-FL125
EGD 202A FL60-FL100
EGD 202B
FL100-FL225
EGD 202C FL100-FL225
D203/23.0
(Outside Hours/18.0)
OCNL/50.0
ABERYSTWYTH
ELAN VALLEY
WINDFARM
WESSEX ASR
COTSWOLD ASR
N864 A FL145+
N864 A FL125+
N864 A FL95+
N864 A FL75+
L9 A FL145+
L9 A FL125+
L9 A FL95+
L9 A FL75+
CARMARTHEN
1⁵
1⁷
2⁸
2⁷
2¹
1⁶
3⁰
3³



High Density RPAS



Air Corridor Network

All Waypoints

ID	Lat. °	Lon. °	Linked
1	55.856	-4.27	2, 21
2	55.856	-4.278	1, 3, 85
3	55.856	-4.302	2, 4
4	55.868	-4.318	3, 5
5	55.868	-4.341	4, 6
6	55.877	-4.363	5, 7
7	55.887	-4.383	6, 8
8	55.892	-4.404	7, 9
9	55.903	-4.422	8, 10
10	55.907	-4.445	9, 11
11	55.924	-4.47	10, 12, 52
12	55.929	-4.497	11, 13, 66
13	55.934	-4.578	12, 14
14	55.962	-4.668	13, 15, 74
15	55.999	-4.715	14

Selected Waypoint

Waypoint #1

Predicted population density at waypoint:

16.0

Traffic Production

Producer? (# aircraft) □ 103.76

Connected waypoints

ID	Lat. °	Lon. °	Link capacity	Link distan...
2	55.856	-4.278	∞	508
21	55.867	-4.271	∞	1207

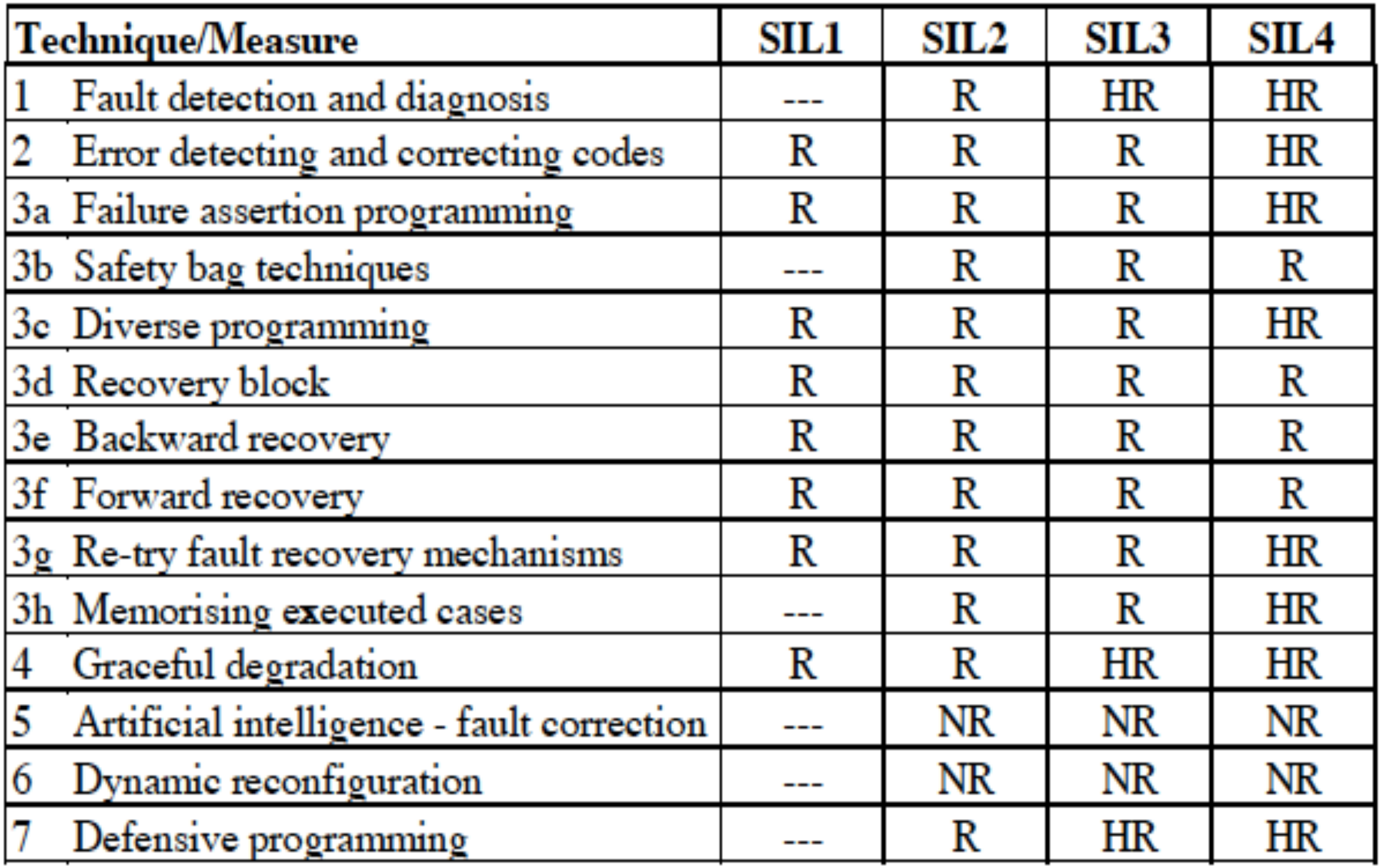
WP103 55.7964°N 4.1895°W
population: 16.00 people/km²
traffic: 97.65 aircraft

Altitude 12 km Lat 55.7964° Lon -4.1895° Elev 201 meters

Federal Aviation Administration

- Notice of Proposed Rulemaking:
  - Certification of Small Unmanned Aircraft Systems (RIN 2120–AJ60).
- RPAS under control of ground pilot equivalent levels 1 & 2.
- Automated control for specific operations providing that the pilot retains ‘line of sight’ with the vehicle; levels 3 and 4.
- Full autonomy banned without specific waivers, restrict ops in experimental zones away from controlled airspace.
- 2015 1,000+ companies had FAA333 exemptions

61508, 26262 and Limits on AI



Technique/Measure	SIL1	SIL2	SIL3	SIL4
1 Fault detection and diagnosis	---	R	HR	HR
2 Error detecting and correcting codes	R	R	R	HR
3a Failure assertion programming	R	R	R	HR
3b Safety bag techniques	---	R	R	R
3c Diverse programming	R	R	R	HR
3d Recovery block	R	R	R	R
3e Backward recovery	R	R	R	R
3f Forward recovery	R	R	R	R
3g Re-try fault recovery mechanisms	R	R	R	HR
3h Memorising executed cases	---	R	R	HR
4 Graceful degradation	R	R	HR	HR
5 Artificial intelligence - fault correction	---	NR	NR	NR
6 Dynamic reconfiguration	---	NR	NR	NR
7 Defensive programming	---	R	HR	HR

Cybernetics and Artificial Intelligence

- Artificial intelligence:
  - Influenced by theories of human cognition;
  - Physiological models - neural networks;
  - Semantic models – formal reasoning.
- Machine learning:
  - More general term than artificial intelligence;
  - not necessarily linked to human cognition;
  - Generalize from training set…
- Eg Fuzzing and genetic algorithms.

Solution 1: Adversarial Approaches

- Manipulate the test set to be really hard.
- How do we define ‘hard’?
  - Traditionally testing insufficient for high SILs.
- Google and others use “the real world”:
  - Ethical issues placing public at risk;
  - How long do you conduct the studies?
  - Risk exposure implies 10^6 hours etc?

Solution 2: Theoretical Verification of ML

- Research topic for neural networks.
  - Show results stable for region of input.

- Huang et al 2017:
  - Scalable verification of multi-layer neural nets;
  - Assumes subset of hidden units in NN relevant;
  - Limits scope of classifier to be considered.

- Limits of region based verification:
  - Cannot imagine all possible inputs;
  - Limits on regions for stability are ad hoc/conservative.

US National Highway Traffic Safety Administration

- *Level 0:* Driver completely controls the vehicle at all times;
- *Level 1:* Individual controls are automated, such as automatic braking;
- *Level 2:* 2+ controls automated, eg adaptive cruise control + lane keeping;
- *Level 3:* Driver can fully cede control of all safety-critical functions in certain conditions. Car senses when conditions require driver to retake control and provides a "sufficiently comfortable transition time” (Tesla S);
- *Level 4:* Vehicle performs all safety-critical functions for the entire trip, with the driver not expected to control the vehicle at any time, including all parking functions. Google lack physical controls.

Google Cars-> Waymo

- Started in 2009, Sebastian Thrun:
  - Costs about $150,000 per vehicle (Lidar).
- Safety performance:
  - 170,000 miles/ month, 125,000 autonomously;
  - Well over 1 million miles;
  - 23 vehicles/14 minor collisions on public roads;
- Only one incident where vehicle to blame;
  - Swerves to avoid sand bags and hits bus.

Waymo in Phoenix, Arizona

- Chrysler Pacific Minivans.
- Massive scale – 100 deployed.
- Twice surface area of San Francisco.
- Part of everyday life…
- Total Waymo test fleet 1000+

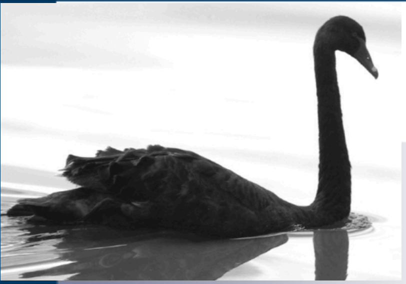
Known Limitations

- Pittsburgh Right and Brussels Left (Priority).
- Cannot use about 99% of US roads.
- Cannot obey temporary road signs.
- Trash, debris, pot holes are big concerns.
- What if humans request you to stop?
  - Most obviously with police officers…

Inconsistent Regulatory Guidance

- Germany:
  - Fed Highway Inst. Auton. vehicles dont meet existing law;
  - Each state grants exemptions ‘if there is a driver in the driver’s seat who has full legal responsibility’.
- France,
  - Testing zones with changes to driver training;
  - Allow ‘large-scale’ tests of self-driving cars/trucks.
- Sweden
  - Volvo ‘Drive Me’ test restricted areas around Gothenburg.

Challenges for Risk Assessment



	Organisational	Individual	Technical
Black Swans	How to manage finite resources to plan for very rare events?	How to mitigate human contribution to risks we never experienced?	How to ensure sufficient range of ‘black swan’ scenarios are considered?
Artificial intelligence	How to show systems that emulate human cognitive behavior acceptably safe?	How to help operators interact with autonomous systems?	How to test non-deterministic autonomous systems?
Cyber security	How much to invest when the risk changes and is uncertain?	How to assess the human contribution to security?	How to protect systems when the past is no predictor of future risks?

University of Glasgow

Cyber Security Risk Assessments

- CRAMM (UK) qualitative risk tool.
- EBIOS (FR) identifies residual risks.
- ISO 13335-2 guidelines for IT security.
- ISO 27005 information security risk management.
- ISO 31000 business risk management.
- IT-Grundschutz (D) Federal IT baseline protection
- MAGERIT (SP) maturity model
- MEHARI harmonized risk, excel support
- Etc.

Cybersecurity Risk

- Amundrud, Aven and Flage (2017):
  - Risk = f(asset_value, threat, vulnerability)
  - Risk = asset x threat x vulnerability
  - Risk = threat x (vulnerability x consequence)
  - Risk = threat x vulnerability x consequence

Organisaitonal: Cyber Risk Assessment

- Threat_Scenario =

  (Attacker, Asset, Method)

- Risk =

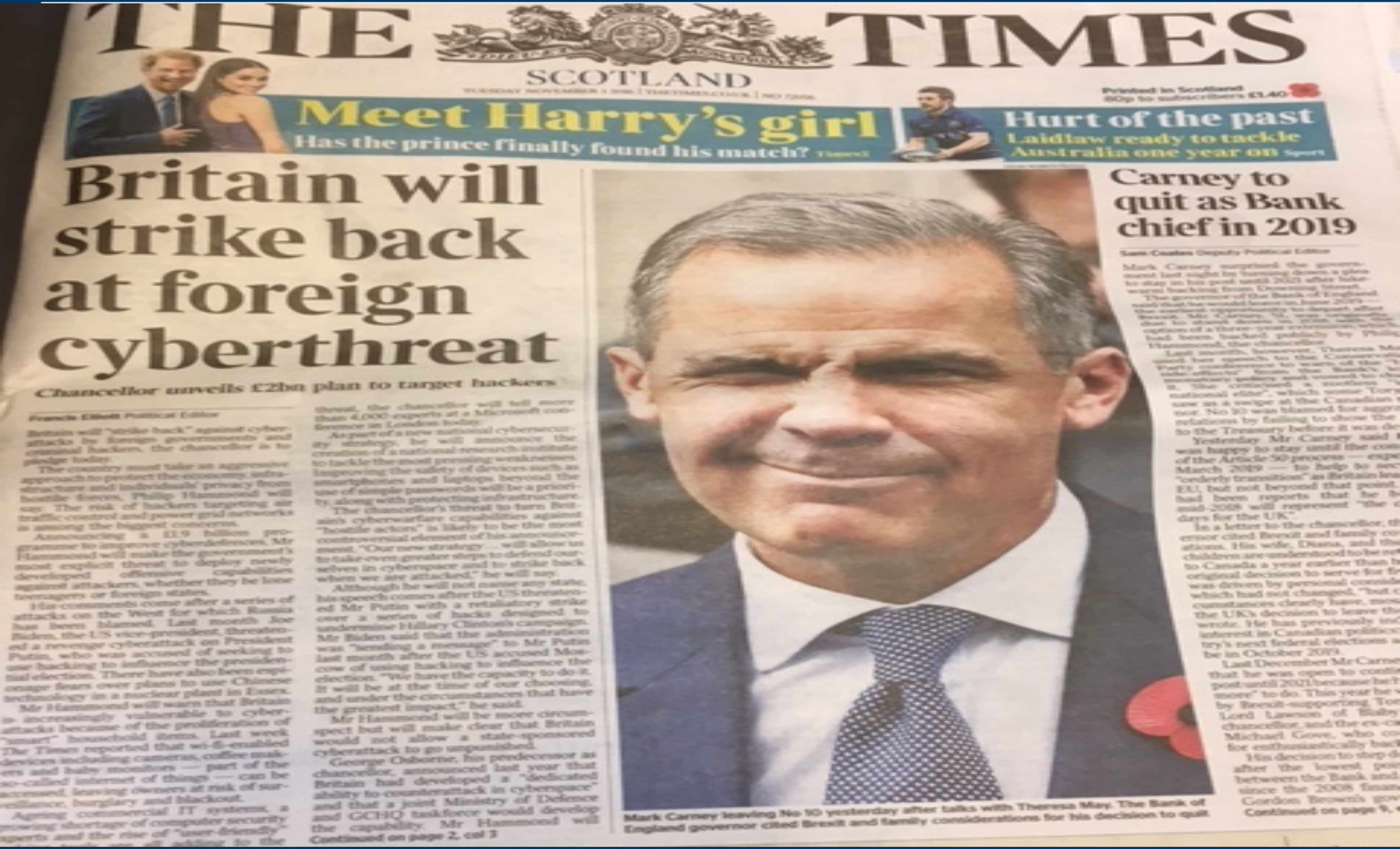
  Probability(Threat_Scenario)

  x Consequence(Threat_Scenario)

Attack Scenarios

- Scenario 1:
  Distributed Denial of Service on Airport's internet connection
- Scenario 2: Deep infiltration to steal data
- Scenario 3: Major integrity loss
- Scenario 4: Blended attack
- Scenario 5: Low Level Attack on APOC ICS infrastructure

The Cyber Arms Race?



THE TIMES
SCOTLAND
TUESDAY NOVEMBER 1 2016 | THETIMES.CO.UK | NO 72056
Printed in Scotland
80p to subscribers £1.40

Meet Harry's girl
Has the prince finally found his match? Times2

Hurt of the past
Laidlaw ready to tackle Australia one year on Sport

Britain will strike back at foreign cyberthreat

Chancellor unveils £2bn plan to target hackers

Francis Elliott Political Editor

Britain will "strike back" against cyber-attacks by foreign governments and criminal hackers, the chancellor is to pledge today.

The country must take an aggressive approach to protect the economy, infrastructure and individuals' privacy from hostile forces, Philip Hammond will say. The risk of hackers targeting air traffic control and power grid networks is among the biggest concerns.

Announcing a £1.9 billion programme to improve cyberdefences, Mr Hammond will make the government's most explicit threat to deploy newly developed offensive capabilities against attackers, whether they be lone teenagers or foreign states.

His comments come after a series of attacks on the West for which Russia has been blamed. Last month Joe Biden, the US vice-president, threatened a revenge cyberattack on President Putin, who was accused of seeking to use hacking to influence the presidential election. There have also been espionage fears over plans to use Chinese technology in a nuclear plant in Essex.

Mr Hammond will warn that Britain is increasingly vulnerable to cyber-attacks because of the proliferation of "smart" household items. Last week The Times reported that wi-fi-enabled devices including cameras, coffee makers and baby monitors — part of the so-called internet of things — can be accessed, leaving owners at risk of surveillance, burglary and blackout.

Ageing commercial IT systems, a growing shortage of computer security experts and the rise of "user-friendly" hacking tools are all adding to the threat, the chancellor will tell more than 4,000 experts at a Microsoft conference in London today.

As part of a new national cybersecurity strategy, he will announce the creation of a national research institute to tackle the most pressing weaknesses. Improving the safety of devices such as smartphones and laptops beyond the use of simple passwords will be a priority, along with protecting infrastructure.

The chancellor's threat to turn Britain's cyberwarfare capabilities against "hostile actors" is likely to be the most controversial element of his announcement. "Our new strategy ... will allow us to take even greater steps to defend ourselves in cyberspace and to strike back when we are attacked," he will say.

Although he will not name any state, his speech comes after the US threatened Mr Putin with a retaliatory strike over a series of hacks designed to undermine Hillary Clinton's campaign. Mr Biden said that the administration was "sending a message" to Mr Putin last month after the US accused Moscow of using hacking to influence the election. "We have the capacity to do it. It will be at the time of our choosing and under the circumstances that have the greatest impact," he said.

Mr Hammond will be more circumspect but will make clear that Britain would not allow a state-sponsored cyberattack to go unpunished.

George Osborne, his predecessor as chancellor, announced last year that Britain had developed a "dedicated ability to counterattack in cyberspace" and that a joint Ministry of Defence and GCHQ taskforce would develop the capability. Mr Hammond will

Continued on page 2, col 3

Mark Carney leaving No 10 yesterday after talks with Theresa May. The Bank of England governor cited Brexit and family considerations for his decision to quit

Carney to quit as Bank chief in 2019

Sam Coates Deputy Political Editor

Mark Carney surprised the government last night by turning down a plea to stay in his post until 2021 after lukewarm backing from Downing Street.

The governor of the Bank of England said that he would leave in June 2019 — the earliest opportunity to depart after Brexit. Mr Carney, 51, was originally due to stand down in 2018, with the option of a three-year extension, which had been backed publicly by Philip Hammond, the chancellor.

Last month, however, Theresa Ma
used her speech to the Conservativ
Party conference to warn of the "b
side-effects" from the Bank's lo
monetary policy and vowed to cha
it. She criticised a rootless "i
national elite", which some Tory
saw as a swipe at the Canadian
not. No 10 was blamed for aggr
relations by failing to show the
to the Treasury before it was de

Yesterday Mr Carney said
was happy to stay until the com
of the Article 50 process — expe
March 2019 — to help to sec
"orderly transition" as Britain lea
EU, but not beyond that point.
had been reports that he b
mid-2018 will represent "the c
days for the UK".

In a letter to the chancellor, th
ernor cited Brexit and family co
ations. His wife, Diana, and the
children are understood to be ret
to Canada a year earlier than h
original decision to serve for fi
was driven by personal conside
which had not changed, "but c
cumstances clearly have, mos
the UK's decision to leave th
wrote. He has previously ind
interest in Canadian politics
try's next federal elections a
be in October 2019.

Last December Mr Carne
that he was open to conti
post until 2021 because he h
more" to do. This year he w
by Brexit-supporting Tor
Lord Lawson of Blaby
chancellor, and the ex-c
Michael Gove, who co
for enthusiastically back

His decision to step d
after the lowest poi
between the Bank and
since the 2008 finan
Gordon Brown's gov

Continued on page 9, c

Uncertainty in Cyber Security

- No confidence in cyber risk assessments:
  - Past does not predict the future (Hume);
  - We cannot trust induction.
- Series of examples relevant to ATM:
  - French bank’s makefile;
  - Chinese hospital patients;
  - Stuxnet/Black energy attack;
  - UK VOIP attack.
- How worried should we be??

Overview

- New uncertainties in Air Traffic Management.
- “Black swan” events seem more common.
- Artificial Intelligence and machine learning.
- Cyber security is under increasing threat.

Challenges for Risk Assessment

	Governmental	Organisational	Individual	Technical
Black Swans	What does ‘acceptably safe’ mean for Black Swan events?	How to manage finite resources to plan for very rare events?	How to mitigate human contribution to risks we never experienced?	How to ensure sufficient range of ‘black swan’ scenarios are considered?
Artificial intelligence	How to promote industry and innovation without exposing society to risk?	How to show systems that emulate human cognitive behavior acceptably safe?	How to help operators interact with autonomous systems?	How to test non-deterministic autonomous systems?
Cyber security	How to protect public and dissuade other nations from attacking?	How much to invest when the risk changes and is uncertain?	How to assess the human contribution to security?	How to protect systems when the past is no predictor of future risks?

Potential Solutions for Risk Assessment

	Governmental	Organisational	Individual	Technical
Black Swans	Regulatory requirements for contingency planning?	Foundations of resillience engineering.	Foundations of resillience engineering.	Common mitigations address multiple scenarios.
Artificial intelligence	Waivers to regulations and segregation to reduce exposure.	Requirements for exhaustive testing and legal reporting framework.	Train humans on modes of interaction with AI systems?	Place bounds on non-determinism, use adversarial scenarios.
Cyber security	NIS Directive and development of offensive weapons.	Simplified rapid risk assessment based on scenarios.	Audit internal provisions, control the supply chain.	Cyber situation awareness (develop offensive techniques)

Overview: take Home Message

- Three “new” concepts/challenges:
  - Black Swans, Artificial Intelligence, Cyber Security.
- One common concern;
  - The Death of Risk Assessment:
- One focus for technical innovation:
  - How do we sustain hazard analysis?
  - How to engineer factors “we know we don’t know”.

Afterword…

- So far we kept it simple.
- Think about the interfaces.
  - AI applied to cyber security (fuzzing);
  - Cyber security of autonomous vehicles;
  - Using Black Swans in cyber weapons.
- How to assess risks of these innovations?

Any Questions?