A View from the Cockpit: Exploring Pilot Reactions to Attacks on Avionic Systems

Matthew Smith*, Martin Strohmeier*†, Jonathan Harman, Vincent Lenders† and Ivan Martinovic*
*Department of Computer Science, University of Oxford, UK,
first.last@cs.ox.ac.uk
†Cyber-Defence Campus, armasuisse Science + Technology, Switzerland
first.last@armasuisse.ch

***Abstract*—Many wireless communications systems found in aircraft lack standard security mechanisms, leaving them vulnerable to attack. With affordable software-defined radios readily available, a novel threat has emerged which allows a wide range of attackers to easily interfere with wireless avionic systems. Whilst these vulnerabilities are known, predicting their ultimate effect is difficult. A major factor in this effect is how flight crew respond, especially whether their extensive training in fault handling helps them to manage attacks.**

To investigate this we conducted a user study, inviting 30 Airbus A320 type-rated pilots to fly simulator scenarios in which they were subjected to attacks on their avionics. We use wireless attacks on three safety-related systems, based on existing literature: Traffic Collision Avoidance System (TCAS), Ground Proximity Warning System (GPWS) and the Instrument Landing System (ILS). To analyze their response, we collected control input data coupled with closed and open interview responses.

We found that all three attack scenarios created significant control impact and disruption through missed approaches, avoidance maneuvers and diversions. They further increased workload, distrust in the affected system, and for each attack, at least a third of our participants switched off the system entirely—even if they were important safety systems. All pilots felt the scenarios were useful, with 93.3% feeling that simulator training for wireless attacks could be valuable.

I. Introduction

Over the past few decades, flying has become ever safer culminating in the year 2017, where not a single death was recorded for commercial passenger air travel [3]. As a whole, the aviation industry and regulators have achieved this with a meticulous focus on safety; for example having stringent requirements on the testing, maintenance and certification of an aircraft.

Two key components of this all-permeating safety mindset are pilot training and on board safety systems. Regular training and assessment of pilots using flight simulator scenarios helps to prepare them in the safe handling of many flight situations. On the other hand, the many wireless technologies onboard an aircraft help to increase situational awareness for pilots and air traffic control (ATC) alike.

Network and Distributed Systems Security (NDSS) Symposium 2020
23-26 February 2020, San Diego, CA, USA
ISBN 1-891562-61-4
https://dx.doi.org/10.14722/ndss.2020.23022
www.ndss-symposium.org

If these avionic systems malfunction or are not used as intended, the consequences can be serious. One example is the case of an inoperative transponder onboard a Delta Airlines aircraft in March 2011, which remained undetected for ten minutes [14]. During this time it flew in close proximity to three aircraft—a working transponder would have helped avoid these unsafe situations. In extreme cases, equipment malfunction can cause loss of life. In 2006, two aircraft collided in Brazil partly due to a failing transponder not relaying collision avoidance messages [1].

Similar to many industries with safety-critical components, aviation is currently working on securing its infrastructure against the new threat of cyber attacks. In this process, all wireless technologies have come under scrutiny, as they almost in their entirety lack fundamental security mechanisms [49]. A subset has been shown to be exploitable under laboratory conditions using widely accessible software-defined radios (SDRs) and software tools (e.g. [10], [43], [46]).

Since the publication of these proof-of-concept demonstrations, the aviation sector has been discussing the severity and reality of this threat, particularly to safety systems. Recent research from the U.S. Department of Homeland Security indicating remote compromise of a Boeing 757 aircraft was dismissed by the manufacturer, who claimed confidence in the security of its aircraft [11]. Several surveys on the perspectives of pilots and other aviation professionals indicate split opinions. Some believe attackers could succeed in creating 'unsafe flight conditions'—the prominent view, however, is that such attacks are mitigated already through aviation's extensive safety systems and culture [2], [48].

Unfortunately, security research into avionics, such as in [10], [43], [46] and [47], has shown that the threat is not addressed by safety-oriented design. Instead, this kind of design deals with random mechanical, electronic, or human failure, rather than deliberate and targeted attempts to subvert the system. As such, an important question is whether attackers can induce failures or unintended behavior using wireless attacks, which then go on to negatively impact the safety of an aircraft.

A standard security assessment approach to this faces a number of challenges. Flight hardware is extremely expensive and difficult to use in isolation, making the construction of a real-world testbed prohibitive for independent research. Furthermore, the flight crew have ultimate authority over how an aircraft is flown, so their response to attacks can do anything

from amplifying the effects to mitigating them entirely. Hence, understanding ow pilots in the loop manage attacks is a necessary requirement to gauge the true impact of wireless attacks, especially on disruption and safety of the aircraft.

To better assess this impact we conducted a user study, simulating three wireless interference attacks in a cockpit environment. Our work recruits 30 professional airline pilots to fly scenarios in a flight simulator during which they are subject to realistic cyber attacks. Some of our attacks are based on related work, whilst others are novel and based on analysis of theoretical vulnerabilities and real-world interference incidents. We consider three heavily used safety-related systems: the Instrument Landing System (ILS), Traffic Collision Avoidance System (TCAS) and the Ground Proximity Warning System (GPWS).

Contributions: This paper identifies how pilots respond to remote wireless attacks on avionics with the intention of highlighting systems requiring further attention. Our contributions are as follows:

- We implement realistic wireless attacks on avionics in a flight simulator for three systems: collision avoidance, instrument landing and ground proximity.
- We run experiments with 30 Airbus A320 pilots to understand how flight crew respond to these attacks.
- We analyze in-simulator and interview debrief results from the experiments, quantifying attack impact on the aircraft as well as lessons learned from the study.

We begin with background in Sec. II, before outlining our threat model in Sec. III. We discuss systems and attacks in Sec. IV, then cover our experimental method in Sec. V. Our results are presented in Sec. VI, followed by discussion in Sec. VII. We discuss lessons learned in Sec. VIII and conclude in Sec. IX.

II. Background

Whilst cyber security in aviation is a more recent concern, investigation into the effectiveness of flight simulators for training is more developed. In this section we consider the background for both of these areas.

A. Cyber Security in Aviation

Increasing awareness of cyber threats in aviation has spurred early-stage research into attacks and countermeasures. An early analysis into the vulnerability of the Automatic Dependent Surveillance-Broadcast (ADS-B) surveillance system generated more widespread attention [10]. At the threat modelling level, several works assess feasible types of attack. In [28], the highlighted threats are spoofing, exploiting, denial of service and counterfeiting. Our study focuses on spoofing and denial of service attacks. In [37], specific threats to aviation security are enumerated, including possible consequences of attacks on collision avoidance systems. We directly assess some of these effects.

Furthermore, technical research into the security of secondary surveillance radar (SSR) systems has assessed the constraints on an attacker aiming to inject, modify or delete SSR messages [43], and provided a thorough assessment of the potential security solutions available [47]. More recently, some other systems such as those used to assist landing were shown to be vulnerable to SDR-based attacks [40]; we discuss this work in more detail in Sec. IV.

Awareness about cyber attacks varies, as demonstrated in [49]. The authors survey aviation professionals on their perceptions on the security of a range of different avionic systems. Whilst there is awareness that the systems are not inherently secure, there does not appear to be significant concern that attacks could affect operational capability.

B. Simulator Training

Time spent in the simulator is a vital part of professional pilot training. A body of research analyzes the configuration of simulator scenarios such that they transfer most easily to flying the real aircraft. Early research indicated that it provides notable benefit over aircraft-only training [24]. However, it is not a given that high-fidelity simulations transfer skills well, and the literature suggests that well-designed scenarios are vital in equipping pilots effectively [39], [12].

One of the key factors in cyber attacks is that there may be no forewarning, leading to surprise and loss of capacity. In [29], a survey of aviation incident reports highlights that 'normal' events can be surprising to pilots when they occur out of context, i.e. alerts when the conditions do not warrant it. The authors in [32] and [7] consider this with respect to stall recovery maneuvers, a regularly tested skill for pilots. Both papers find that pilots struggled to follow even well-known procedures when the stall occurred in unexpected conditions.

Addressing this, the authors of [33] argue that unpredictability and variability in simulator training improves performance when encountering surprise scenarios. While their work uses failure scenarios instead of malicious interference, the arguments remain valid.

C. Simulating Cyber Attacks

Some work addressing simulation for cyber security has begun to emerge. In [23] the authors conduct a human factors focused study to assess how pilots respond to an attack on ground-based navigation systems. They find that pilots under attack lose some monitoring capacity, and that warnings can help mitigate this. The authors of [6] (and the extended [15]) conduct a more avionics-focused set of attacks, looking at six variants of navigation and flight management system threats. Multiple attacks inserted over the course of one flight with the intention being to assess if pilots notice the attacks. They found that most attacks were identified during flight, however some happened without detection.

Our work differs in that we focus specifically on systems that are either entirely or partly safety-critical with the attacker instead aiming to cause disruption as a minimum. We also chose to explore a different set of systems and cover the principles of these attacks in technical detail.

III. Threat Model

We presume a moderately resourced attacker, able to buy commercially off-the-shelf antennae, amplifiers, and SDRs. A high-end set up could cost under $15,000, including a

scientific-level SDR (e.g. the Ettus USRP at around $6000), high-power amplifiers for VHF and UHF (likely to be bespoke or scientific hardware but approximately $7000 in total) and directional VHF and UHF antennas, one for each attack (approximately $500-$1000 depending on the number required).

For lesser-equipped attackers, a more basic set up could be achieved for under $3000 by using an SDR such as the HackRF ($300), a commodity amplifier ($1000) and omnidirectional antenna (around $200 for both). This approach does have limitations though; it would not have sufficient power to carry out attacks over long distances, so would be less effective.

We presume that the attackers in this paper have a high-end set up. This would enable them to transmit at sufficient power to communicate with airborne aircraft. We also presume they have the capability to develop software, or use existing open-source tools, to interfere with aircraft systems. Our attacker can deploy their systems remotely or create a mobile platform from which to do this.

Threat Actors: We consider three threat actors: activists, terrorists and nation states. Activists intend to cause disruption to raise the profile of their cause, usually with low resource but high levels of personnel. On the other hand, terrorists aim to disrupt or destroy with the intent of creating a chilling effect or fear. They can be moderately resourced and are unlikely to care about collateral damage. Most extreme is the nation state who primarily intend to disrupt in order to paralyze infrastructure. They are well resourced and are likely to be concerned about attribution and collateral.

Attack Aims: We focus on attempts to cause disruption, rather than destructive impact. This is due to our work looking to the fundamentals of the attacks, in which disruption is likely to be the first effect, though these may have destructive variants. This can include diversions to alternative airports, excessive movement away from planned routes or *go-arounds*, i.e. a missed approach to land followed by a second attempt. Our work intends to identify how crew handle these attacks as a base case, which is indicative of the impact they would have under a stronger threat model. As such, we believe that based on our results, future work could focus on destructive attacks. We discuss this further in Sec. VII.

Furthermore, we are careful to ensure the experiment is fair. In scenarios where the aircraft is put at risk of crashing it would be unrealistic to assess pilot response outside of their normal environment. For example, we could not accurately assess response times if controls are slightly different to a full simulator or real aircraft. We cover our experimental setup and its limitations in Sec. V.

IV. Systems and Attacks

We now concisely describe the systems and the attacks used in the experiment, including the expected crew responses. More detailed technical descriptions of the attacks can be found in App. A.

A. Ground Proximity Warning System

A fundamental part of an aircraft's 'safety net', the Ground Proximity Warning System (GPWS) provides early warning of the aircraft becoming too close to terrain [5].

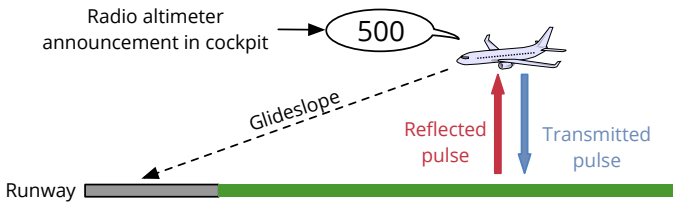


Fig. 1: Normal system operation with radio altimeter determining height above ground.

1) System Description: Two versions of this system exist—the original GPWS, and the newer Enhanced GPWS (EGPWS) which incorporates GPS and a terrain database. The subsystem used in this study is the same in both. Taking a range of sensor inputs, GPWS provides alarms of situations leading to collision with terrain [51]. It has a range of alert modes; we focus on excessive closure on terrain, or Mode 2 [5]. Mode 2 GPWS uses a radio altimeter to determine of the height above ground level (AGL) and the rate of closure on nearby terrain; one of the primary uses is on approach to landing. We provide a representation of this on approach in Fig. 1.

2) Attack Description: Our attacker creates a spurious ground proximity alert when the aircraft is close to landing (i.e. on final approach) to negatively impact situational awareness and cause an unwarranted go-around. As a result, aircraft will then have to perform a second approach or divert to a different airport. During this time, the aircraft will be using extra fuel, incurring delay and increasing workload for the pilots. By transmitting specifically-crafted false radar pulses on final approach, the attacker causes GPWS to believe that the terrain closure rate is significantly higher than in reality. This will trigger a *'Terrain Terrain, Pull Up'* alarm, even though the aircraft is close to the ground yet within a 'safe' range. In terms of resource, the timing of this attack is likely to present the greatest challenge and will need appropriate software developed. Hardware requirements are simpler, as it requires the directional transmission of a pulse on a given frequency; this aspect is similar to a DEF CON 2019 talk which attacked police speed radars [50]. As such, this attack can be carried out by lower capability and resourced threats.

3) Expected Response: Whilst the response will depend on the aircraft and airline, there are common principles [45]. In most conditions we expect a terrain avoidance maneuver on alarm, i.e. a steep climb to a safe altitude. In our scenario, this will lead to a missed approach. However, below 1000 ft above aerodrome level (AAL), with full certainty of position, crew can choose to not follow this. Due to the surprise element, we expect the typical response to be a missed approach. On following approaches we expect participants to have identified unexpected behavior and disregard the warnings.

4) Simulator Implementation: We simulate the attack by triggering the GPWS '*Terrain, Terrain, Pull Up*' alarm starting at 500 ft AGL on approach to Runway 33 at Birmingham, increasing by 250 ft for each subsequent attack. This emulates the ability of an attacker to add some unpredictability to the attack. One of the limitations of this approach is that the point at which the attack actually triggers can vary between 450 ft and 500 ft AGL, and the radio altimeter display does not show

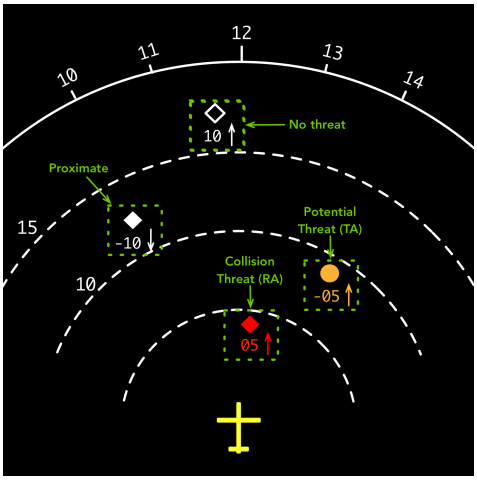


Fig. 2: Representation of TCAS display to pilots based on the Airbus A320/330 cockpit Navigational Display (ND).

the change caused by the attack on the cockpit display but just announces the alarm.

B. Traffic Collision Avoidance System

Although ATC manage airspace with high precision, aircraft may end up closer than is safe. This *loss of separation* can, in the worst case, result in a mid-air collision. One example, mentioned above, occurred in March 2011 where a Delta aircraft took off with an inactive transponder. This was the first in a line of errors which resulted in it becoming too close to three other aircraft before resolving the issue [14]. If active, Traffic Collision Avoidance System (TCAS) provides a technical means to avoid this, and has been mandated on aircraft with more than 30 seats since 1993 [17], [25].

1) System Description: TCAS makes use of an aircraft's transponders to interrogate nearby aircraft [17]. Analyzing the responses to these interrogations allows the object aircraft to calculate whether those aircraft will become too close [18].

Based on lateral and vertical proximity to nearby aircraft, visual representation and alerts are given to crew similar to that in Fig. 2. These come in two steps; first is a *traffic advisory* (TA), in which the traffic is typically displayed to the pilot as amber and an aural alert of 'traffic' is given. If the nearby aircraft becomes closer, a *resolution advisory* (RA) is given. An RA will contain specific instructions for the flight crew, i.e., to climb or descend at a given rate, or hold vertical speed. These instructions are decided between the two aircraft automatically to deconflict the situation. RA instructions must be followed within seconds.

In the cockpit, crew have control over the alerting level by selecting *Standby*, *TA-ONLY*, or *TA/RA*. For most of a flight, TCAS will be set to TA/RA in which full alerting is provided. TA-ONLY does not issue RAs, whereas Standby performs no TCAS interrogations or conflict resolution [19].

2) Attack Description: In our scenario, the attacker aims to cause crew responses to TCAS by triggering TAs and RAs despite no aircraft being nearby. This is intended to burn unnecessary fuel, break from air traffic control clearances, affect situational awareness, and cause knock-on alerts for other aircraft. This may result in diversions or switching TCAS off. To achieve this, the attacker generates TCAS responses for a false intruder aircraft, which approaches the object aircraft until it reaches the alert regions. We refer to the attacked aircraft as *target* and the injected aircraft as *false*. Some of the technical capabilities required for such an attack can be seen in [4], wherein the author explores how to trick a target aircraft to track an attacker-generated aircraft. This is a powerful attack requiring expensive equipment and the ability to cover a large geographic area; the threat actor most likely to be capable of this is a nation state. However, an attack covering a smaller region could be carried out by a less well-equipped attacker, e.g. a terrorist group.

3) Expected Response: As following an RA is compulsory, we expect pilots will comply with at least the first instance, follow the instructed maneuver [44]. From there, we expect some participants to doubt RAs and eventually turn the alert level down from TA/RA to TA-Only or Standby. On average, we expect participants to follow 3-4 RAs before reducing the alert level or switching the system off.

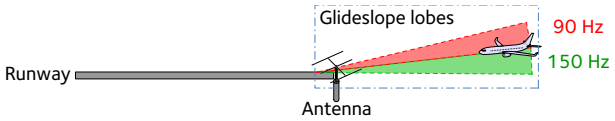
4) Simulator Implementation: Within the simulator, we enact a strong attacker who covers a large geographic area, attempting to trigger 10 alerts over the course of the flight. We varied the angle and speed of approach by the false aircraft. Each participant had the same sequence of false aircraft approaching in the same way. False aircraft began to be injected when the target aircraft flew above 2000 ft, after which the first injection began. If the participant chose to turn the TCAS sensitivity to TA-Only, they would still receive TAs but not RAs. This attack was undertaken by using an invisible aircraft model which travelled towards the target aircraft. Further work would improve the realism of this, such as more realistic flight patterns to avoid tipping off participants to the attack.

C. Instrument Landing System

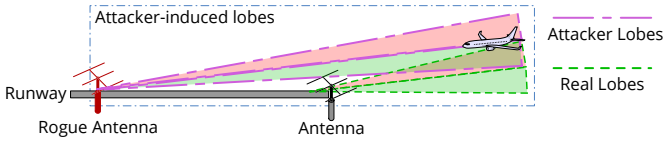
The Instrument Landing System (ILS) allows precision landings even in poor weather conditions. Since aircraft must follow specific arrival routes into an airport, ILS is an important part of managing pilot workload and is the default approach type for most airports. In extreme cases, ILS allows aircraft to automatically land at sufficiently equipped airfields.

1) System Description: ILS consists of two components: localizer (LOC) and glideslope (GS) [21]. A localizer provides lateral guidance and alignment, centered on the runway centerline, whereas the GS provides vertical guidance to a touchdown zone on the runway. Typically, the GS will provide a 3° approach path, though this depends on the specific approach and airport [41]. It is supplemented by Distance Measuring Equipment (DME), which provides the direct distance to a beacon without directionality.

Transmission powers of the GS and LOC are 5 W and 100 W respectively [21]. On the carrier frequencies for the GS and LOC, overlapping 90 Hz and 150 Hz lobes provide guidance with the overlap forming the centerline on the approach path. The aircraft uses the relative strength of these lobes to identify where it is with respect to the optimal GS



(a) Glideslope under normal operation.



(b) Glideslope under attack with rogue antenna. Note how the aircraft touchdown zone is now at the far end of the runway. This means that if the glideslope is followed to touchdown, there may not be enough runway to slow down.

Fig. 3: Representation of normal and under-attack glideslope operation, based on diagrams from [21].

Fig. 4: Picture of experimental setup.

and centerline of the runway. A diagram of a GS can be seen in Fig. 3a.

GSs and LOCs are monitored for accuracy to at least 10 nmi beyond the runway, as well as being protected from interference to 25 nmi [42], [41]. It is important to note that 'protection from interference' here means avoiding systems using nearby frequencies, rather than malicious interference.

Separately, approach lighting provides an out-of-band check for crew on approach—Precision Approach Path Indicators (PAPIs) are configured to match to the angle of the GS. When an aircraft is on the correct GS, the PAPIs will show two red and two white lights; otherwise more red or white lights are shown as appropriate [22].

2) Attack Description: Here, the attacker is aiming to cause unnecessary missed approaches as a result of a tampered GS, similar to that in [40]. In turn, this will use additional fuel, introduce delay and potentially force aircraft to divert to a different airport. A secondary aim might be to force crew to use a different, also attacked, approach method.

The attacker replicates the real GS but with the touchdown zone short or long of the legitimate touchdown zone by transmitting a replica signals from aside the runway. Since they will not be able to station themselves on the runway, they will operate outside the airfield perimeter. This somewhat matches the legitimate GS signal which is transmitted aside the runway to avoid aircraft clipping the antennae.

Crucially the signals would be the same as a real GS, so would not be identifiable by a high rate of descent, as common GS issues can be. The difference induced by the attacker would be subtle. For a typical 3° GS, moving the touchdown zone 1 km along the runway creates a consistent height difference between the real and false GS of approximately 52 m, or 172 ft. This could fall within a margin of error on approach, especially whilst further away from the runway.

Such an attack is moderately difficult due to creating the correct signal and transmitting from an appropriate position. We consider all of our threat actors capable of this but the attack success may depend on the capability and equipment.

3) Expected Response: Since this attack will see the false GS track slightly above the real GS, it is unlikely to be immediately obvious that it is incorrect. We expect most participants to follow the GS until they are below cloud at around 1000 ft, at which point they will notice a continued slight discrepancy in AGL according to approach charts. They may also notice such a discrepancy using the PAPIs, as they will show four white lights. At this point, we expect them to be between 500–1000 ft AGL and opt for a missed approach and go-around.

4) Simulator Implementation: In the simulator, an attacker transmits a false GS at the far end of the runway with an effective shift of 2.05 km, or 1.27 miles, creating a difference between the false and true GS of 107 m, or 352 ft. Due to the way in which ILS is implemented in the simulator software, we could not replicate also having a 'real' GS. To account for this, we operated on an assumption that the attacker transmits at a higher power than the real GS in an effort to force capture on to the false GS. The manipulation remains in place regardless of how many approaches are made. We treat the participant aircraft as if it is the first to encounter the attack, with ATC not observing previous aircraft having difficulties.

V. Experimental Method

Since our attacks were specifically designed to examine responses, we wanted to allow participants to react in real time. To do this we used a flight simulator, partially recreating a cockpit environment—in this section, we describe the experimental setup used. The work was approved by our local ethics committee with reference number R54139/001.

A. Participants

We recruited 30 pilots who had current Airbus A320 type-rating or had held it in the past few years but had since moved to larger Airbus aircraft. Our sample was recruited through pilot forums, and open to pilots of any level of experience, First Officer or Captain. This is appropriate since pilots are trained and kept current with a homogeneous skill set for a given type of aircraft. Thus, all pilots are similarly skills-equipped to handle the scenarios we presented to them. Participants were compensated for their time with a gift voucher.

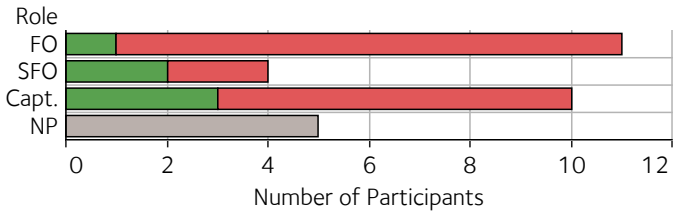


Fig. 5: Participant role demographics in flight crew: First Officer (FO), Senior FO (SFO) and Captain (Capt.). NP is where participants chose not to provide data. Green bars indicate a training role, red for those without and grey is 'not provided'.

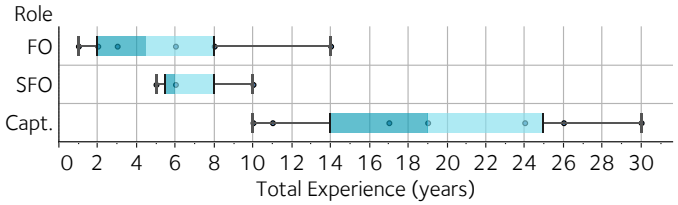


Fig. 6: Plot of participant commercial flying experience by role: First Officer (FO), Senior FO (SFO) and Captain (Capt.).

1) Demographics: We collected demographics from participants with an option not to provide information if desired. In Fig. 5 we show participants by both role and whether they hold a trainer role. We split into the three key crew roles in order of increasing seniority: First Officer, Senior First Officer and Captain. Furthermore, the colors indicate whether the participant trains other pilots as part of their job.

In Fig. 6 we provide a chart of participant commercial flying experience, grouped by role. Note that captains have a wide range of years of flying experience due to the requirements for taking a captain role varying between company and location. The median total years of commercial flying experience for a Captain was 19, for an SFO was 6 and for an FO was 4.5.

B. Protocol

For the purposes of control, we used the same weather conditions, traffic, and route for four runs. Pilots were asked to fly between two international airports, cruising at 12,000 ft, for a total flight time of around 30 minutes. Since the setup was single-pilot, the experimenter provided support in enabling modes, pressing buttons or selecting cockpit views for the pilot. These actions were done solely at the command of the pilot and the experimenter provided no decision support. Additionally, the experimenter provided ATC information to each pilot where relevant, such as approving clearances to change altitude.

Each pilot was given the first run as a familiarization flight, in which they could get used to the controls of the simulator. The following three runs included some form of attack with each followed by a short debrief interview. At the end of the third attack and debrief, we asked some questions on the study as a whole. We used the same order of attacks for each participant.

TABLE I: Summary of participant actions and responses to debrief yes/no questions. For some participants, the question was not applicable due to previous actions such as landing regardless of alarm, hence N/A. Percentages are of all participants, for each question.

		Response					
		Yes		No		N/A	
Attack	Question	#	%	#	%	#	%
GS	Q5–Trust	1	3.3	25	83.4	4	13.3
	Q6–Safety	19	63.3	11	36.7	-	-
	Q7–Same	28	93.3	2	6.7	-	-
TCAS	Q5–Trust	4	13.3	22	73.4	4	13.3
	Q6–Safety	28	93.3	2	6.7	-	-
	Q7–Same	30	100.0	0	0.0	-	-
GPWS	Q5–Trust	0	0.0	12	40.0	18	60.0
	Q6–Safety	14	46.7	16	53.3	-	-
	Q7–Same	27	90.0	3	10.0	-	-

The interview assessed the pilot response to each attack, focusing on perception of impact, trust, workload and safety. This was done with closed questions, but we allowed the participants to provide additional comments if they wished. Only data from closed questions were used in our numerical analysis. Interview questions are outlined in Sec. VI and provided in full in App. B. We recorded data from the simulator to correlate with interview responses. This included control inputs, aircraft position, speed and heading. The details of the attacks were explained by the experimenter in debrief.

We note that our study has some limitations, discussed further in Sec. VII. At this point we note two limitations. First, whilst participants knew that they were taking part in a study looking at cyber attacks on avionic systems, they did not know about the timing or type of attack. Also, since the interview was conducted by the experimenter, we acknowledge that this may bias results to be more positive than if we had conducted this anonymously. This is mostly relevant to interview questions on the effectiveness of this approach as training, and we note this where appropriate.

C. Equipment

Our hardware consisted of two high-end gaming PCs, running X-Plane 11 and an aftermarket Airbus A330 model as no reliable A320 models were available, seen in Figure 4 [31]. We checked the model fidelity with type-rated Airbus A320 pilots to ensure sufficient similarity to an A320. We provided non-type-specific hardware controls, since the majority of flying on such an airliner involves manipulating automatic flight, rather than directly flying with manual controls. Participant opinions on the equipment are presented in Sec. VII-E.

VI. Results

We now discuss the data collected from simulator scenarios and participant interviews. Interview response data can be seen in Tab. I and Fig. 7, with full data for this figure provided in App. C. Responses are on the following scales:

- *Q1.* **Confidence** in the response being the correct one, on a scale from 1, *very confident*, to 5, *very unconfident*.

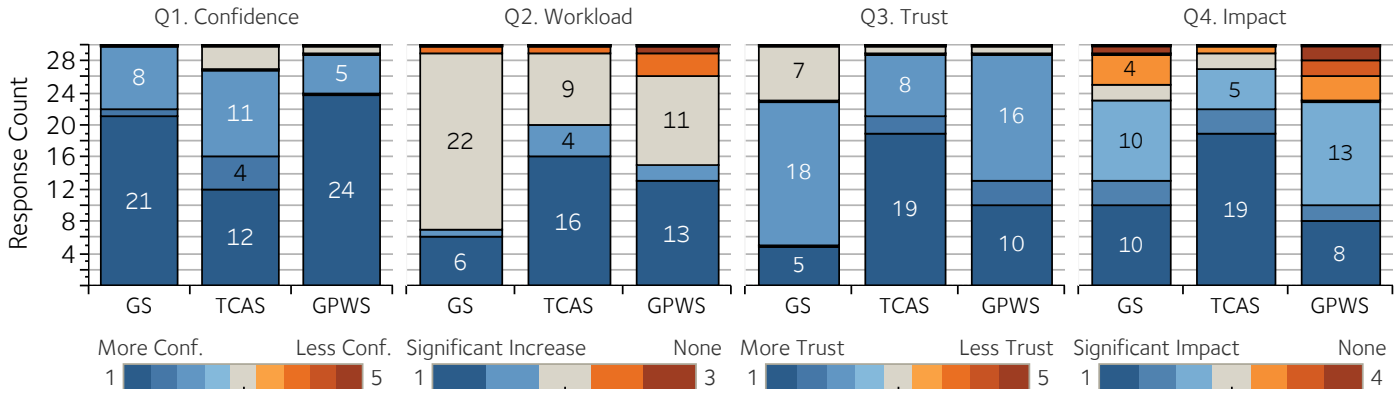


Fig. 7: Stacked bar charts for participant scale responses on Q1–4. Orange represents the most 'negative' responses, i.e. no effect, with blue 'positive', i.e. significant effect. Tabular data is provided in Tab. VI.

- *Q2.* **Workload** due to the attack, on a scale from 1, *no increase*, to 3, *significant increase*.
- *Q3.* **Trust** in systems affect due to the attack, on a scale from 1, *much more trust*, to 5, *much distrust*.[1]
- *Q4.* **Impact** on the flight due to the attack, on a scale from 1, *significant impact*, to 4, *no impact*.

The Q2 and Q4 scales differed from scoring 1–5 to better represent their topic. For Q2, we are measuring any increase from baseline workload hence the scales are half of those in Q1 and Q3. In Q4 we again measured from a baseline of *no impact* but aimed to collect a more granular response coupled with qualitative answers.

We also recorded *yes/no* responses for the following:

- *Q5.* Whether they would trust systems under attack later in flight, N/A if they did not respond to the attack.
- *Q6.* If participants felt the attack put the aircraft in a less safe situation.
- *Q7.* If participants would respond the same way in a real aircraft (i.e. free of simulation restrictions).

Tab. I summarizes the response to these; note that the table designates some responses to Q5 as not applicable in cases where actions preclude the question. In the case of GPWS, this is N/A is when the participant switches the system off, for TCAS it is when they did not change the system mode away from TA/RA and for GS it is when a participant landed on the first approach despite the attack.

A. GPWS Attack

First, we look at the GPWS scenario. We assess participants primarily on their actions, i.e. go-around, land with the alarm sounding or switch GPWS off and land, before considering their scale responses.

Response: Participants generally responded as expected, split between those opting for a terrain avoidance maneuver (thus a missed approach) and those disregarding the warning in order to land. Vertical profiles for all participants on the first approach are plotted in Fig. 8, with Tab. II, showing participant actions split into landing, aborting the approach (i.e. *go-around*) and turning the system off. If a participant lands or turns the system off they 'complete' the flight on that approach so are not included in subsequent approaches, e.g. the 10 pilots who land in approach one are then no longer

[1]In this study we consider *temporary* trust, i.e. trust during the scenario. We cannot assess longer term trust as we did not carry out repeated simulations for each attack, per participant.

TABLE II: Action taken during GPWS attack. If a participant lands, they are not included in the numbers of following approach. Percentages are of participants in that approach.

		Action Count		
Approach	Action	#	%	# Participants
1	Land	10	33.3	30
	Go-around	20	66.7	
2	Turn off	11	55.0	20
	Land	8	40.0	
	Go-around	1	5.0	
3	Turn off	1	100.0	1

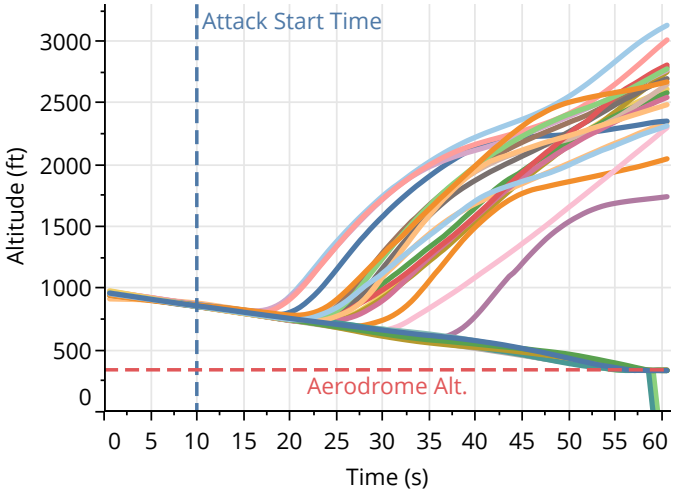


Fig. 8: Plot of time against altitude for first approach under GPWS attack. Each line is a participant. Eight land and disregard the alarm, on account of being sure of their position.

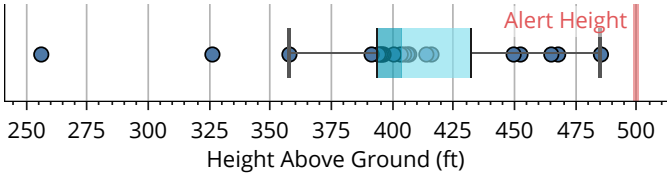


Fig. 9: Minimum heights reached by participants opting to go around in first approach of the GPWS attack.

flying so are not included in approach two statistics.

Two thirds of participants went around on the first approach as a result of the alarm; these participants remarked that their choice was an automatic one due to procedure. This shows that triggering such an attack can cause arbitrary go-arounds with reasonable chance of success, at least on the first approach.[2]

In Fig. 9 we plot the minimum altitude reached during the first approach for those who did not go on to land, i.e. those who performed a go-around. In this chart, a higher altitude indicates a faster decision to abort the approach. Across participants, we found that the go-around began at height $\bar{x}$: 403.9 ft (s: 51.1 ft). Some outliers in the form of later responses do exist, also visible in Fig. 8. Most participants responded within 100 ft of the alarm with an interquartile range of 29.7 ft. The relatively small interquartile range around 100 ft after the attack triggered—with most outliers sitting towards higher altitude—shows that the participants are well-drilled in responding to this alarm. From an attacker's point of view this lessens the safety impact of the attack as a decision to abort the approach is quickly reached.

As indicated earlier, we expected most pilots to follow the terrain warning and execute a well-drilled maneuver, not allowing the aircraft to become unsafe due to low altitude. However, of the 10 who chose not to go around on the first approach, seven identified the alarm as spurious and so were happy to disregard it. Five of these participants felt they could do this due to good weather, implying that poor visibility would push them to abort.

To manage the attack, 11 participants switched the system off prior to a second approach, finding the alarm distracting. An attack causing GPWS to be switched off has the potential for further erosion of safety—indeed, of the 12 who switched it off, none said that they would trust the system later in the flight. This is of benefit to the attacker in some situations. In the case presented as part of our scenario, the terrain was broadly flat and some pilots had flown into the airport before. In less familiar circumstances or in challenging terrain (e.g. mountainous), pushing crew into a situation which forces them to switch a safety net off could result in unsafe situations. However, this is counteracted somewhat by the pilots mostly feeling that either the system was giving spurious alarms so was distracting anyway. On top of this, most of the participants who did switch the system off commented that they were sufficiently sure of their location relative to the terrain that they did not see it as an undue risk.

[2]In one instance, the attack triggered late; however, in debrief, the participant noted that they would have taken the same course of action and landed regardless.

Perception: As seen in Tab. I, 14 (46.7%) participants felt that this attack put the aircraft in a less safe situation. The numbers are lower compared to other attacks as the response is in itself a safety maneuver, though some pilots felt that due to the extreme nature, the aircraft is at additional risk. This is because the maneuver involves high engine power and a steep climb, possibly into the vicinity of other aircraft.

Fig. 7 shows that this scenario has the least impact as assessed by the participants—even so, it was judged to have 'some impact' on average, with 8 (26.7%) saying it was 'significant'. For workload, there was on average 'some increase' with 13 (43.3%) feeling there was a 'significant increase'. These results imply that the attack is much more of a nuisance than a risk; the pilots identify and manage the issue quickly before it can become more serious but still have to deal with extra work. On top of this, a number of remarks were made about the startle factor involved on what appeared to be a normal approach. This can significantly impact workload as the event is far outside the expected set of possible events.

An inevitable consequence of pilots identifying spurious warnings was that trust in the system was eroded during the scenario. From Tab. I we can see that 12 (40.0%) participants would not trust the system later in flight, with the other 18 participants expressing their distrust by switching it off (hence N/A). Matching this to Fig. 7, 29 (96.7%) participants felt at least 'some distrust' towards the system after the attack.

Generally, confidence in response was very high, with an average score of 'very confident'. The majority of participants (27, 90%) said they would take the same course of action in a real aircraft. This is likely to be due to terrain alarms being such a high priority—with a high risk if the decision is wrong—that pilots are trained to quickly respond in a particular way with minimal scope for choice. Indeed, those who said that they would behave differently in a real aircraft suggested that they would have opted for a missed approach rather than landing. In turn, this means that the vast majority of participants would choose to abort the first approach. To some extent, this confidence can be exploited by an attacker. Knowing that on balance, pilots are likely to abort a landing if a terrain alarm is triggered, they can be relatively sure that such an attack will at least cause a nuisance in the first instance.

Evaluation: The reaction was relatively consistent seemingly due to strict procedure on how to handle terrain alarms. We can deduce that:

- An attacker has a good chance of triggering an unnecessary missed approaches on the first approach by an aircraft, due to startling, temporary workload increase and prescribed reaction to terrain alarms,
- The attack lacks longevity—pilots quickly identified the alarm as spurious and disabled or ignored it,
- A safety reduction occurs but is limited, only becoming worse in unfavorable conditions.

As a result, the attack does not cause a significant reduction in safety and has short-term disruption potential but is easily managed. Considering that this is the most technically simple attack, it could be attractive to lower capability threat actors seeking disruption alone, such as activists.

B. TCAS Attack

Next, we consider the TCAS attack. Results indicate that this is the most concerning attack to the participants.

Response: An action summary is given in Tab. III. We provide the 'end-state' of the selected TCAS mode (e.g. if a participant selects *TA-Only* then *Standby*, they are under *Standby*) against actions taken which fall outside of normal flight actions. Actions are categorized into *continue on route*, i.e. no extra action taken, *avoidance maneuver*, in which the participant changes course beyond responding to an RA such as changing flight level, or *divert to origin*, i.e. return to the departure airport. Some 26 participants (86.7%) turned TCAS to TA-Only during flight, with 11 (36.7%) switching it to Standby thus turning the system off. Participants switched to TA-Only after $\bar{x}$: 4.5 RAs (s: 1.7), then down to standby after another $\bar{x}$: 2.8 TAs (s: 2.1). Two participants went straight from TA/RA to Standby, one after three RAs, another after six. With the exception of one participant, all followed the instructions of the RA whilst TCAS was set to TA/RA mode. This meant following collision avoidance maneuvers involving steep climbs and descents.

These actions highlight two problems created by this attack. First, the attacker has the opportunity to push the aircraft away from 'normal' flight by triggering alerts which cannot be ignored. The range of outcomes to these alerts, coupled with the fact that the average pilot complied with over four RAs before reducing sensitivity, shows that there is no straightforward response. This indicates that the attack has a confusion factor.

Next, the attack caused the majority of participants to reduce the sensitivity of TCAS and in some cases, switch it off completely. This is a loss of situational awareness which could allow unsafe situations to develop later in flight. Many participants stated that this response was a trade-off between the additional workload of responding to TAs and RAs if the system is left on against the loss of full use of TCAS if it is switched off. They also felt that the additional workload was too great. Furthermore, some participants noted that the distraction brought about by repeatedly responding to the alerts meant they had less time to deal with other aspects of flight.

Looking at the control response in more detail, three of those eventually turning the transponder to TA-Only and three of those turning it to Standby took avoiding action. The action itself varied per participant but for some involved climbing above the planned cruise altitude or making horizontal maneuvers to try to avoid the attacker's false traffic. On top of being unpleasant for passengers, this increases the risk of incursion into the path of other aircraft; particularly dangerous when TCAS is apparently malfunctioning. Furthermore, two participants diverted back to the origin airport rather than continue with malfunctioning TCAS, which would incur significant costs (discussed in Sec. VII). Three of the remaining participants felt that TCAS was providing spurious returns but felt the risk of downgrading the system to TA only was too high and instead opted to follow the RAs as issued. Such a response would lead to an uncomfortable flight, excessive fuel use from repeated climbing and the possibility of becoming too close to other aircraft. The final participant was not aware of the ability to go to TA-Only in the simulator and so remained in TA/RA.

TABLE III: Responses to the TCAS attack scenario, mapping the final selected TCAS mode against actions or maneuvers taken by the pilot. Percentages are of all participants.

	Final Selected TCAS Mode							
	TA/RA		TA-Only		Standby		Total	
Action	#	%	#	%	#	%	#	%
Continue on route	4	13.3	10	33.3	8	26.7	22	73.3
Avoidance maneuver	0	0.0	3	10.0	3	10.0	6	20.0
Divert to origin	0	0.0	2	6.7	0	0.0	2	6.7
Total	4	13.3	15	50.0	11	36.7	30	100.0

These responses are important as even though the false aircraft generated by the attacker were identified by most participants as spurious, they still caused a range of emergency actions. This indicates that the attacker has a significant amount of influence though this attack.

Perception: Looking to Fig. 7 and Tab. VI, we can see that 27 (90.0%) pilots felt that the attack had at least 'some impact', with 19 (63.3%) feeling that it had 'significant impact'. In comparison to the other attacks, this judged to be the most impactful by far. Coupled with the vast majority of participants identifying that the TCAS returns were spurious, a variety of reasons were provided such as unusual intruder behavior, frequency of RA or that ATC were not observing the intruders. Some participants commented that they experience one RA a year at most during their job, so seeing multiple, rapid RAs was a sign of unusual activity. A further 29 (96.7%) participants felt that there was at least 'some increase' in workload, typically due to having to respond to regular RAs and dealing with periodic distraction. An unduly increased workload creates further problems for the crew managing the situation and can lead to errors.

Considering perceived safety, 28 (93.3%) pilots felt that the attack put the aircraft in an unsafe—or potentially unsafe—situation. A variety of reasons were provided by participants with three themes emerging: effect on other aircraft, crew or passenger injury and distraction (as discussed above). The first presents a unique risk to this attack. Since responding to TCAS RAs results in the aircraft making an emergency maneuver, this can result in other aircraft nearby losing separation and thus being issued with TCAS alerts. This might cause the attack to trigger a chain of alerts, disrupting every aircraft involved. In this situation, the outcome becomes less predictable since multiple aircraft are involved, each reacting in their own way. The other cause for concern was for those onboard who might be moving about the cabin, thus injured in an extreme maneuver such as an RA. This is especially true of RAs triggered at higher altitudes where passengers and cabin crew may not be sat down with seatbelts on.

Similarly, 29 (96.7%) participants felt they had at least 'some distrust' in TCAS during the scenario. Again, this is problematic as it indicates that an attacker with moderate ability can sow distrust in critical aircraft safety systems during flight. One participant described this as a "crying wolf" effect,

wherein TCAS was being triggered so often that they might start to disbelieve it even though they cannot refuse to act.

Evaluation: In this scenario, the most common option was to reduce the alerting level of TCAS to either only notify of traffic (TA-Only) or to switch the system off. We also identified some common outcomes:

- Repeated, unexpected alarms cause pilots to make a choice: deal with the disruption and distraction, or turn the system off and possibly lose the safety benefit,
- Time taken to identify the attack was longer than the other attacks, indicating a confusion factor,
- No prevailing way to handle the attack emerged, with pilots split between a range of actions.

Although this attack is the most difficult to carry out, our results suggest that it has the greatest impact on the crew, aircraft, surrounding traffic and the passengers.

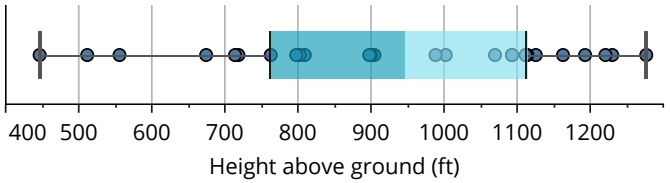
C. Glideslope Spoof

The final attack is the glideslope spoof, where an attacker aims to capture a pilot on a false GS and cause missed approaches. We focus on the first approach, in which the participants knew least about the attack; our results also indicate that most pilots identify a problem on this approach.

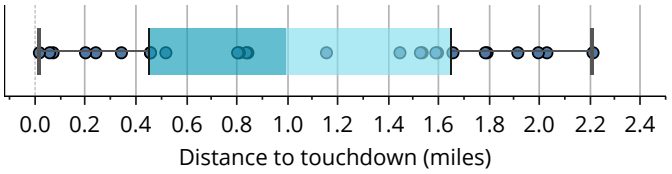
Response: On encountering the attack, 4 (13.3%) participants chose to land anyway on account of having a good visual picture. This means that they identified a problem—that the aircraft was too high compared to the real glideslope—but felt that weather conditions and terrain were good enough to correct course and land anyway. Of the 26 (86.7%) participants choosing to abort the first approach, three aborted their second approach too but landed on their third. Participants aborting approaches identified a problem but felt they needed to go around either to use a different type of approach or to allow more time to diagnose the issue. The choices for subsequent approaches were as follows:

- 1 (3.3%) used a VHF Omnidirectional Range approach,
- 2 (6.7%) used a Surveillance Radar Approach (SRA), which relies on higher involvement with ATC,
- 8 (26.7%) flew a localizer only approach (LOC DME) on account of identifying GS problems,
- 9 (30.0%) avoided ILS completely, and used an Area Navigation (RNAV) approach, which is based on GPS,
- 6 (20.0%) flew a visual approach (i.e. no landing aids) due to good conditions.

This split highlights that the attack invokes a response gray area and creates unpredictability. Eleven participants chose to forgo ILS completely and use SRA or RNAV approaches as they could not identify the issue precisely. However, eight were happy to use LOC DME, relying on the localizer component of ILS, since they felt that they had identified that just the GS was affected. Such a range of responses also indicates that the attack is only likely to be effective for the first or second approach as after this most participants avoid the glideslope. Even so, by this point the aircraft has already been disrupted through a go-around which will cause delay and use extra fuel.



(a) Height above ground level at point of first go-around.



(b) Distance from runway touchdown zone at first go-around.

Fig. 10: Box plots of participants performing a go-around on the first approach under the glideslope attack.

Fig. 10a and 10b show box plots for the height above ground level (in feet) and distance (in miles) from the runway touchdown zone, respectively, for when participants opted to abort the first approach. These charts demonstrate the wide range of 'abort points' observed during the attack, indicating flight crew confusion or startle caused by the false GS. Many participants noted that it was hard to identify the issue quickly as initially, everything seemed fine; it was only as the approach continued and the PAPIs became visible that the problem was more obvious.

Looking to the average case of the go-around for the 26 pilots aborting their first approach, altitude was at $\bar{x}$: 930.0 ft (s: 235.8 ft), for distance to touchdown at the abort point $\bar{x}$: 1.1 miles (s: 0.7 miles). Since preparation for a go-around takes a few seconds, the average case abort decision takes place slightly before the above altitude and distance, i.e. just as participants descended below 1000 ft.

Considering that a 3° GS has a rate of descent of 700 ft/min, this means the go-around begins with just over a minute to touchdown. In poor weather, this might be the first time the pilots see the runway with only a short amount of time to abort the approach. That the attack is subtle enough for the aircraft to get so close to landing demonstrates how difficult it is to clearly identify that an ILS attack is under way. In this case, forcing a late go-around would inevitably impinge on safety.

Perception: As shown in Fig. 7, 13 (43.3%) participants found the attack had 'some' impact or greater; less than the TCAS attack but judged more significant than the GPWS attack. This appears to be due to the attack being relatively easy to manage once diagnosed—though the diagnosis took some time—with some participants noting that faulty glideslopes are experienced in practice. Furthermore, a range of routine options exist outside of ILS, unlike TCAS and GPWS.

The GS attack had a small workload increase with 22 (73.3%) participants claiming 'some' increase, the lowest average score of the three. This may be due to the GS attack developing gradually at a higher altitude with PAPIs providing visual reference for correctness, which the other attacks did not have. This allows more time for participants to respond, in

contrast to TCAS or GPWS attacks which trigger alarms and need immediate attention.

Despite this, 19 (63.3%) participants felt that the attack made the aircraft less safe. A number of participants noted that this attack would be harder to deal with in other situations. In worse weather conditions such as extremely low visibility they would have fewer reference points against which they could check the glideslope. This would make it hard to even identify that an issue exists until very late in the approach. Some participants also commented that if the glideslope was short, rather than long, of the runway threshold (i.e. touchdown was before the runway started), it would be significantly more dangerous. This is because the approach might look normal until very late at which point the aircraft would be at risk of landing off the runway.

Although there was little additional workload, 26 (86.7%) participants performed at least one go-around as a result of being unsure about the approach, instead seeking the safest option possible. Here, this involved taking a second approach, in many cases with a different landing system. Some pilots noted that low fuel situations would limit the options and possibly only allow one more approach, making the attack more difficult to manage.

As with TCAS and GPWS, the attack caused 'some' distrust in aircraft systems, with 23 (76.7%) participants remarking 'some' or 'significant' impact. However, some participants correctly identified that the ground systems were at fault and so did not distrust the aircraft. In this situation, they were able to diagnose the issue and 'cut out' ILS, thus mitigating the attack. As such, attackers would have to consider other vectors if they wished to guarantee disruption. This is supported by responses to Q5, on trusting the system later. In Tab. I we can see that of the 26 (86.7%) participants, who did perform a go-around, all but one would not trust the GS on a second approach.

Evaluation: Generally, this attack was considered a nuisance rather than a significant safety issue but did manage to disrupt. Our results indicate that:

- Whilst the attack consistently caused first approach disruption, its effect was limited beyond this as participants used other approach methods,
- The subtlety of the attack and a lack of alarms meant that aircraft got close to the runway—within a couple of minutes before landing—before they had to abort,
- After an initial problem diagnosis, the attack was fairly easily managed with little excess workload,
- Variants such as poor weather may be much more difficult to handle and pose a greater safety risk.

Despite the limited effect, the attack can cause short term disruption through triggering go-arounds, in turn burning excess fuel and increasing delay. However, it is likely that attacking consecutive aircraft would see ATC instructing aircraft not to use ILS. Whilst a more sophisticated attacker might tamper with multiple systems, this significantly increases cost and the risk of detection.

TABLE IV: Summary of attack costs, equipment requirements, difficulty and overall impact. H is *high*, M is *medium* and L is *low*.

	Attack Name		
Actor	GS	TCAS	GPWS
Cost	M	H	L
Equipment	M	H	M
Difficulty	M	H	L
Disruption/Impact	M	H	L
Safety Effect	M	H	L

TABLE V: Mapping of attacks against threat actors.

	Attack Name		
Actor	GS	TCAS	GPWS
Activists	✓	✗	✓
Terrorists	✓	✗/✓	✓
Nation State	✓	✓	✓

VII. Discussion

We now discuss results across attacks and within wider contexts such as cost and compared to system faults.

A. Attack Comparison

In Tab. IV we provide a high-level summary of the costs, equipment requirements and difficulty based on Sec. IV, followed by the potential for disruption/impact and safety effect derived directly from participant impact (Q1), and safety (Q6) assessment in Sec VI.[3] Using this, we then map the attacks to threat actors in Tab. V.

Whilst all aircraft were handled safely in our experiment, there appears to be a meaningful effect on safety from the TCAS attack, with variants on the GS attack also able to create unsafe situations. Despite participants taking the safest option in the circumstances, TCAS saw 93.3% participants feel that the attack made the aircraft less safe with GS at 63.3%. For TCAS, this could be the uncertainty of the situation, with pilots not expecting false alarms; for GS, the safety concern comes from how late the discrepancy is apparent and the situation this leaves the aircraft in. GPWS safety was split with 46.7% feeling that the aircraft was less safe. This is due to the terrain avoidance maneuver being the de facto safe option, making the automatic response the safest response. However, the GPWS attack highlights the interplay of safety and security; even though most pilots took the safe option, they still felt they were compromised by factors out of their control.

All three attacks have the potential to cause some degree of disruption, but the level and means vary. Clearly, TCAS has the greatest potential with the attack causing participants to respond for a longer portion of flight than the others. This is in contrast to GS and GPWS both of which caused an initial disruption but were then quickly managed, with most participants landing by their second approach. However, TCAS

[3]For example, impact an average impact response of 'significant impact' is a score of 'high'. With safety, we base the score on the proportion of respondents judging there to be a safety impact.

is also the most complex attack to carry out, requiring high skill and resource levels. A more simplistic version may be achievable by less capable, highly determined attackers such as terrorist groups. For less capable attackers, the less disruptive attacks such as GS and GPWS are in scope, but as shown, the effects are short-term and identified faster.

By comparing key findings across the three scenarios, we can extract some general insights:

1) **Whilst alarms force action they are quickly turned off or ignored if considered spurious.** In the case of TCAS and GPWS, the procedural need to respond to alarms meant that participants looked at ways to 'manage' this which sometimes involved turning the system to a lower sensitivity level or off. Since these systems are all key to safety, having to switch them off because of their susceptibility to attack is suboptimal.
2) **Attackers can force pilots away from systems.** Best demonstrated in the GS scenario, attacking systems makes participants treat them as faulty and seek to use others. This can lead to further disruption but limits the long-term effect of attacks.
3) **Gray areas can be managed using existing procedure but variability is high.** Whilst safety was compromised by some attacks, all participants handled them without major incident. However, the eventual response and steps to get there varied—significant in some scenarios—partly due to difficulty in diagnosing the problem. This could be exploited by an attacker to create confusion.

The lack of security for the systems in our scenarios not only allows attackers to cause disruption directly but can also mean that the systems become unusable and so are switched off. Considering that pilots are taught to trust cockpit systems and rely on them being accurate, this is a dangerous combination.

B. Comparison to System Faults

Many faults on an aircraft are identified and reported by on board computers then presented to flight crew through screens, warning lights or alarms. Extensive development and testing of the aircraft allows potential faults to be identified and management methods to be provided to crew, usually through checklists. This means that crew are prepared for faulty behavior, usually with a predefined series of actions to take for the safest outcome.

Our scenarios take advantage of edge cases in procedure or develop in ways which do not trigger alarms. Whilst they might have similarities to faults—and are handled in this way by most participants—this can be a confusion factor. For example, in the GS attack, participants noted the slow development of the attack with no other warnings. In the case of TCAS, whilst alarms were going off, participants commented that no checklist exists for spurious TCAS, which led them to eventually turn the sensitivity down as the best decision in the circumstances. Because of this slight difference, although existing training helps pilots to ultimately handle the issue, it might not help them diagnose the problem in the first place.

C. Additional Impact Factors

As discussed in Sec. VI, participants highlighted a number of other factors which would affect the impact of attacks. Weather conditions were prominent; all scenarios would be more difficult to handle in poor visibility. Some participants noted it would be hard to identify the GS attack under automatic landing conditions (i.e. poor visibility), leaving much less time for pilots to respond. Other contributing factors include tiredness and terrain. In response to the GPWS attack, one participant who chose not to go around commented that their action in a real aircraft would depend on tiredness, as well as weather and how busy the crew were. Again in the GPWS attack, others identified that terrain surrounding the airport affects their choice—they would be much more likely to abort an approach in challenging terrain, and less if they are familiar with the airport.

D. Cost of Disruption

We have demonstrated the ability for these attacks to cause missed approaches and diversions. With this in mind, we can estimate the resulting costs.

For go-arounds, as caused by GS or GPWS, we can calculate the cost of a missed approach using a representative Boeing aircraft.[4] For a smaller 737-800 aircraft, the missed approach uses 127 kg (41.79 gal) more fuel than a successful one; for the larger 777-200, it is 399 kg (111.55 gal) more [38]. Coupled with a nominal jet fuel cost of 184.58 c/gal, this costs approximately \$77 for the 737 or \$205 for the 777.[5] Added to the expense of further time in the air—more difficult to predict as it depends on factors such as the airfield and traffic—plus a second approach, which costs approximately \$139 (using 230 kg, or 75.68 gal) or for the 737, or \$516 for the 777 (using 850 kg, or 279.69 gal), this becomes expensive for the airline.

All three attacks created the possibility of having to divert, with four participants choosing to follow this through during the TCAS scenario. Diversions add further expense on top of excess fuel burn, as well as having knock-on effects for scheduling or causing passenger inconvenience. The UK Civil Aviation Authority estimates that these can cost an airline between £10,000–£80,000, depending on the size of the aircraft and location of diversion [9]. For example, passenger disruption causing diversion aboard a Norwegian flight cost €100,000 in 2018 [13], [16]. Closed airports are similarly costly, with drones closing London Gatwick for two days in December 2018 and costing airline Easyjet £15 million [30].

E. Simulation for Training

To assess whether responses were realistic, we asked each participant whether their response to each scenario would be the same in a real aircraft. We found that for:

- **GPWS**, 27 (90.0%) would do the same, and the remaining three would go around in the same scenario again,
- **TCAS**, 30 (100.0%) would do the same,

[4]This is chosen due to the public availability of fuel usage information about Boeing aircraft.

[5]Calculated using IATA Jet Fuel Price Monitor for 18^{th} January 2019 [26].

- **Glideslope**, 28 (93.3%) would do the same with the remaining two opting to go around and revert to RNAV.

We asked each participant for their views the value of such experiments or training in preparation for cyber attack. All participants felt the scenarios were useful, and 28 (93.3%) commented that training for cyber attacks using a simulator would be valuable.

The results suggest that this method can be valuable both in identifying crew response to attacks and providing cyber attack readiness. Furthermore, the fact that the scenarios in this paper lie in procedural gray areas and do not have a series of steps to resolve them provides an ideal opportunity for training. One point of caution is negative training, with some participants noting that care must be taken to avoid training pilots to ignore or distrust their systems.

Finding a balance between awareness and negative training is important to fully prepare pilots for attack scenarios. Currently, pilots are trained to handle a wide range of aircraft faults from diagnosis through to remedy or mitigation. The capability to address these faults is reassessed regularly as part of pilot license revalidation—in the case of commercial pilots, this is usually once or twice a year, often in a flight simulator.

One way to approach this balance would be to include attack simulations in training and revalidation based on known-possible effects, which could be derived from penetration testing or reports of real incidents. Importantly, this would need to be coupled with a comparison to existing faults, how an attack differs and an honest discussion of the likelihood of such an attack occurring. Since pilots are already used to the fact that faults can occur at any time, this simply augments their knowledge with fault diagnosis-style tools for attacks instead.

Ultimately, there should be little difference in how a fault and an attack is handled on the flight deck as both impinge on the function of the aircraft. As our results show, existing fault handling procedure often gets pilots part-way to managing attacks, so additional training can extend these procedures to cover cases where attack effects deviate from failures. We would expect further research to establish such best-practice procedures, with input from both the computer security and the aviation communities.

F. Experimental Limitations

As addressed in Sec. V, there are some limitations to our experimental approach such as not being a full crew complement or taking place inside a full replica cockpit. We acknowledge that this may have some effect on the results and so surveyed participants about it, asking if they felt limited by the simulation set up (Q23, App. B), with 8 feeling heavily limited, 18 somewhat limited and 4 not limited. The average response was 'somewhat', with the main limits being the lack of a second crew member and the general (rather than Airbus specific) controls. We note that these figures are subject to some bias due to the experimenter interviewing the participant.

Prior Knowledge: Since we did not have existing access to pilots to sample, we had to recruit externally. Our recruitment material revealed that the experiment related to aviation cyber security in general but no further details such as the systems being attacked. We felt that this level of prior information was important in recruiting participants as attending an experimental session required a reasonably high level of effort on their part, namely in arranging around a busy flying schedule and usually long-distance travel to our lab.

According to methodology research on human participant studies, we consider our participants naive since they are aware of the topic but not its methods or expectations [36]. Relevant literature suggests that having fully non-naive participants can affect results slightly—one study asked participants to complete a series of tasks twice with some time gap, with up to a 25% reduction in effect [8]. However, a meta-study identified works which suggest that non-naive participants can also be less likely to conform to experimenter expectations [34]. Since our participants only had knowledge of the topic but no specifics, we are confident that participants did not lose naivety. Even if such an effect is significant, we expect that participants were more likely to anticipate malfunction and so be more alert, providing a 'best case' reaction.

VIII. Lessons Learned

Having considered the results of our study, we now look at lessons arising from it, applicable to aviation and transport or infrastructure security scenarios with humans in the loop.

a) Diagnosis is key: Our results show that it is unrealistic to rely on humans to plug the gap between safety and security. Pilots are extensively trained to deal with the many faults which can emerge when flying an aircraft, and this was reflected in the results. However, the attacks generated situations which shared some features with faults but largely were different; they lacked indication of failure. This meant that even though they knew something was wrong, a lot of time had to be spent diagnosing the issue. One way to improve on this would be to factor attack scenarios into existing simulator training schedules, and to add failures caused by attack into existing fault diagnosis and handling procedures. As well as general preparedness, this might help to reduce startle should an attack occur. Furthermore, being upfront with crew about the effects and likelihood of attacks will help them handle said attacks better if they happen.

b) Value of simulation: We uncovered a number of factors which affect how an attack develops that would have been out-of-scope if we had focused on individual components. By taking a wider system view with a simulator, we could allow scenarios to unfold, providing more information about how pilots responded to the attacks. We could also gather information about other factors affecting the response which we might not have considered in our initial analysis, such as typical system behavior or other air traffic. These factors are important in assessing the true impact an attack has—for example, with GPWS, our paper-based analysis indicated that it would be more problematic than it turned out to be. More generally, this approach is especially valuable in systems where humans play a key role and are used to simulators as part of their training; examples include transport such as trains, or nuclear power plant operators.

c) Real usage matters: One of the key motivators of this work is an attempt to understand whether what operators *should* do during an attack differs to what *actually* happens.

Aviation is one amongst many areas of infrastructure known for strict safety rules and robust policies. In theory, systems in these domains should be predictable under attack. However, we found that quirks and oddities of such complex systems can initially mask attacks; in our case, pilots were willing to deviate from strict procedure in order to manage workload or distraction. These were reasoned decisions with the intention to maintain or improve safety, but often ran contrary to what the rules or regulations say. For instance, TCAS is considered to be an extremely important safety system and misuse has led to crashes—in our work, attacking it resulted in it being so distracting that pilots felt they had to turn it off. This led to a contradiction; in theory, the aircraft was less safe as it had a key safety system turned off, however in practice the crew felt they were maintaining safe flight by removing a distraction. As mentioned above, even where outcomes were predicted on paper, real responses varied, in some cases significantly. This is particularly relevant to the wider industrial security and safety-critical system community—when we are thinking about security in complex systems, human operators effectively become an amplifier of effect and understanding how they actually behave is vital.

IX. Conclusion

In this paper, we consider the effects of three wireless interference attacks on avionic systems with respect to the flight crew. We implemented the attacks in a flight simulator and tested how 30 commercial pilots responded. Our results show that all of the attacks have at least some potential for disruption, which in turn could lead to a reduction in safety, financial loss or reputational harm. Crucially, participants often had to make a choice between reducing distraction and turning off key systems, or keeping said systems on; in over a third of cases, safety systems were switched off.

Our results identify the attack on TCAS as the most concerning since it combines widespread inconvenience and potential safety reduction. Both GS/ILS and GPWS also pose problems, though are easier to mitigate on the flight deck. Finally, we conclude that flight simulation for wireless attack awareness or training has potential to aid and prepare crew. Since preventative security by design will not be deployable in the near-term, such training could be highly valuable.

Acknowledgements

The authors would like to thank Jeremy Thomson for his help in testing the simulator and scenarios during the development stage of our experiment.

References

[1] Final Report A-00XCENIPA2008. Technical report, Aeronautical Accident Investigation and Prevention Center, September 2006.

[2] Airline Pilots Association. Aviation Cyber Security: The Pilot's perspective. Technical report, Air Line Pilots Association Int'l, Washington, 2017.

[3] BBC. 2017 safest year for air travel as fatalities fall. https://www.bbc.com/news/business-42538053, January 2018. Accessed on 2018-11-20.

[4] Paul Martin Berges. Exploring the Vulnerabilities of Traffic Collision Avoidance Systems (TCAS) Through Software Defined Radio (SDR) Exploitation. Master's thesis, Virginia Tech, 2019.

[5] Barry C. Breen. *Digital Avionics Handbook*, chapter 21, pages 21.1–21.12. CRC Press, 3rd edition, 2015. gpws chapter.

[6] Jan-Philipp Buch, Robert Manuel Geister, Luca Canzian, Giovanni Gamba, and Oscar Pozzobon. What the Hack Happened to the Flight Deck: Analyzing the Impact of Cyberattacks on Commercial Flight Crews. In *AIAA SciTech 2019*, January 2019.

[7] Stephen M Casner, Richard W Geven, and Kent T Williams. The effectiveness of airline pilot training for abnormal events. *Human factors*, 55(3):477–485, 2013.

[8] Jesse Chandler, Gabriele Paolacci, Eyal Peer, Pam Mueller, and Kate A. Ratliff. Using non-naive participants can reduce effect sizes. *Psychological Science*, 26(7):1131–1139, 2015. PMID: 26063440.

[9] Civil Aviation Authority. Disruptive Passengers. https://www.caa.co.uk/Passengers/On-board/Disruptive-passengers/, 2018. Accessed on 2018-11-20.

[10] Andrei Costin and Aurélien Francillon. Ghost in the Air(Traffic): On insecurity of ADS-B protocol and practical attacks on ADS-B devices. In *Black Hat USA*, pages 1–10, jul 2012.

[11] Joseph Cox. US Government Probes Airplane Vulnerabilities, Says Airline Hack Is 'Only a Matter of Time'. https://motherboard.vice.com/en_us/article/d3kwzx/documents-us-government-hacking-planes-dhs, June 2018. Accessed on 2019-02-15.

[12] N Dahlstrom, Sidney Dekker, R Van Winsen, and J Nyce. Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science*, 10(4):305–314, 2009.

[13] Brendan Dorsey. Hawaiian Airlines Passenger Fined $100,000 for Bad Behavior. https://thepointsguy.com/2017/08/hawaiian-airlines-passenger-fined/, August 2017. Accessed on 2018-11-20.

[14] Eurocontrol. Flying without a transponder—10 minutes is all it can take. *NetAlert*, (19):5, May 2014.

[15] European Aviation Safety Agency. Impact Assessment of Cybersecurity Threats. Technical Report EASA_REP_RESEA_2016_1, European Union, 2018.

[16] Irish Examiner. Limerick court fines man €1,000 after disrupting flight to tune of €100k. https://www.irishexaminer.com/breakingnews/ireland/limerick-court-fines-man-1000-after-disrupting-flight-to-tune-of-100k-674943.html, April 2015. Accessed on 2018-11-20.

[17] Federal Aviation Administration. *Introduction to TCAS II Version 7.1*, chapter 1, pages 5–10. U.S. Department of Transport, 2011.

[18] Federal Aviation Administration. *Introduction to TCAS II Version 7.1*, chapter 1, page 17. U.S. Department of Transport, 2011.

[19] Federal Aviation Administration. *Introduction to TCAS II Version 7.1*, chapter 1, pages 22–24. U.S. Department of Transport, 2011.

[20] Federal Aviation Administration. *Introduction to TCAS II Version 7.1*, chapter 1, pages 17–19. U.S. Department of Transport, 2011.

[21] Federal Aviation Administration. *Instrument Flying Handbook*, chapter 9, pages 9.35–9.38. Number FAA-H-8083-15B. U.S. Department of Transport, 2012.

[22] Federal Aviation Administration. Lighting Systems – Precision Approach Path Indicators (PAPI). https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/lsg/papi/, June 2015. Accessed on 2018-11-23.

[23] Patrick Gontar, Hendrik Homans, Michelle Rostalski, Julia Behrend, Frédéric Dehais, and Klaus Bengler. Are pilots prepared for a cyber-attack? A human factors approach to the experimental evaluation of pilots' behavior. *Journal of Air Transport Management*, vol. 69:pp. 26–37, June 2018.

[24] Robert T. Hays, John W. Jacobs, Carolyn Prince, and Eduardo Salas. Flight simulator training effectiveness: A meta-analysis. *Military Psychology*, 4(2):63–74, 1992.

[25] Steve Henely. *Digital Avionics Handbook*, chapter 22, pages 22.1–21. CRC Press, 3rd edition, 2015. TCAS Chapter.

[26] International Air Transport Association (IATA). Jet Fuel Price Monitor. https://www.iata.org/publications/economics/fuel-monitor/Pages/index.aspx, November 2018. Accessed on 2018-11-20.

[27] International Civil Aviation Organization. *Annex 10 to the Convention on International Civil Aviation—Aeronuatical Telecommunications*, volume 1, chapter 3, pages 3.19–3.20. 2006.

[28] Daniel P Johnson. Civil Aviation and CyberSecurity. http://sites.nationalacademies.org/cs/groups/depssite/documents/webpage/deps_084768.pdf, 2013. Accessed on 2019-02-04.

[29] Janeen A Kochan, Eyal G Breiter, and Florian Jentsch. Surprise and unexpectedness in flying: Database reviews and analyses. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 48, pages 335–339. SAGE Publications Sage CA: Los Angeles, CA, 2004.

[30] Julia Kollewe and Gwyn Topham. EasyJet says Gatwick drone chaos cost it £15m. https://www.theguardian.com/business/2019/jan/22/easyjet-gatwick-drone-cost-brexit-flights, January 2019. Accessed on 2019-02-15.

[31] Laminar Research. X-plane 11. https://www.x-plane.com/, August 2018. Accessed on 2018-11-21.

[32] Annemarie Landman, Eric L Groen, MM Van Paassen, Adelbert W Bronkhorst, and Max Mulder. The influence of surprise on upset recovery performance in airline pilots. *The International Journal of Aerospace Psychology*, 27(1-2):2–14, 2017.

[33] Annemarie Landman, Peter van Oorschot, M. M. (Rene) van Paassen, Eric L. Groen, Adelbert W. Bronkhorst, and Max Mulder. Training pilots for unexpected events: A simulator study on the advantage of unpredictable and variable scenarios. *Human Factors*, 60(6):793–805, 2018.

[34] Austin Lee Nichols and John E. Edlund. Practicing what we preach (and sometimes study): Methodological issues in experimental laboratory research. *Review of General Psychology*, 19(2):191–202, 2015.

[35] Ofcom. UK Amateur Radio License – Terms, Conditions and Limitations. https://www.ofcom.org.uk/__data/assets/pdf_file/0027/62991/amateur-terms.pdf. Accessed on 2018-08-31.

[36] Oxford Reference. Naive participant. "https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100221655". Accessed on 2019-06-06.

[37] Raju Patel. Managing Cybersecurity Risk. https://www.omg.org/news/meetings/tc/va-17/special-events/cybersecurity-pdf/Dr-Raju-Patel-Managing-Cybersecurity-Risk-in-Weapons-Systems-3-21-17.pdf, 2017. Accessed on 2019-02-03.

[38] William Roberson and James A. Johns. Fuel Conservation Strategies: Descent and Approach. *AERO*, (38):25–28, 2010.

[39] Eduardo Salas, Clint A Bowers, and Lori Rhodenizer. It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology*, 8(3):197–208, 1998.

[40] Harshad Sathaye, Domien Schepers, Aanjhan Ranganathan, and Guevara Noubir. Wireless attacks on aircraft instrument landing systems. In *28th USENIX Security Symposium*, Aug 2019.

[41] Dorothy Saul-Pooley. *Radio Navigation & Instrument Flying*, chapter 15, pages 325–329. Pooley's, 2017. Glideslope.

[42] Dorothy Saul-Pooley. *Radio Navigation & Instrument Flying*, chapter 15, pages 325–321. Pooley's, 2017. Localiser.

[43] Matthias Schäfer, Vincent Lenders, and Ivan Martinovic. Experimental analysis of attacks on next generation air traffic communication. In *International Conference on Applied Cryptography and Network Security*, pages 253–271. Springer, 2013.

[44] SKYbrary. Airborne Collision Avoidance System (ACAS). https://www.skybrary.aero/index.php/Airborne_Collision_Avoidance_System_(ACAS)#Complying_with_RAs, 2017. Accessed on 2018-08-30.

[45] SKYbrary. Response to a "PULL UP" Warning. https://www.skybrary.aero/index.php/Response_to_a_%22PULL_UP%22_Warning, 2017. Accessed on 2018-08-30.

[46] Matthew Smith, Daniel Moser, Martin Strohmeier, Vincent Lenders, and Ivan Martinovic. Economy Class Crypto: Exploring Weak Cipher Usage in Avionic Communications via ACARS. In *21st International Conference on Financial Cryptography and Data Security*, Malta, 2017.

[47] Martin Strohmeier, Vincent Lenders, and Ivan Martinovic. On the Security of the Automatic Dependent Surveillance-Broadcast Protocol. *IEEE Communications Surveys & Tutorials*, 17(2):1066–1087, 2015.

[48] Martin Strohmeier, Anna K Niedbala, Matthias Schäfer, Vincent Lenders, and Ivan Martinovic. Surveying Aviation Professionals on the Security of the Air Traffic Control System. In *International Workshop on Cyber Security for Intelligent Transportation Systems (CSITS)*, September 2018.

[49] Martin Strohmeier, Matthias Schäfer, Rui Pinheiro, Vincent Lenders, and Ivan Martinovic. On Perception and Reality in Wireless Air Traffic Communication Security. *IEEE Transactions on Intelligent Transportation Systems*, 18(6):1338–1357, June 2017.

[50] Bill Swearingen. HAKC THE POLICE. Online, August 2019. Accessed on 2020-01-14.

[51] UK Civil Aviation Authority. Ground Proximity Warning Systems. https://publicapps.caa.co.uk/docs/33/CASPEC14.PDF, 1976. Accessed on 2018-12-16.

APPENDIX A
TECHNICAL DETAILS ON THE ATTACKS

A. GPWS

A radio altimeter is a Frequency-Modulated Continuous Wave (FMCW) radar, transmitting pulses on frequency sweeps between 4200 and 4400 MHz. It uses the frequency shift and round-trip time for the received signal to calculate the height above terrain, also referred to as *above ground level* (AGL). Its operation is illustrated in Fig. 11, where Δt is the round-trip time, and Δf is the frequency shift.

The attack aims to replicate the rapid closing of ground by transmitting a ramp of frequencies between 4200 MHz and 4400 MHz. The gradient of this ramp is crafted to incrementally reduce the round-trip time per frequency shift for the signal, creating the illusion of the ground approaching rapidly.

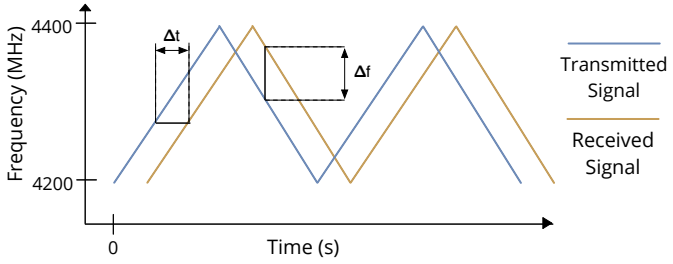
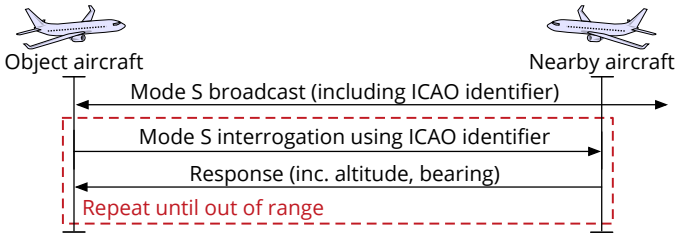


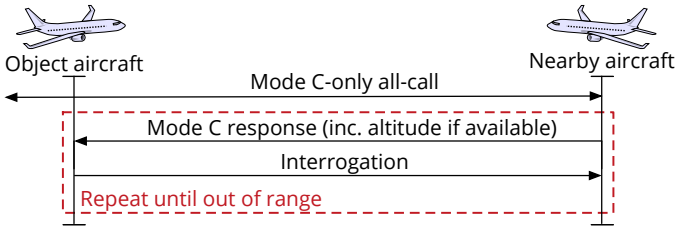
Fig. 11: Frequency-modulated continuous wave (FMCW) radar operation at a static height, for both the transmitted signal and received, reflected signal.

This requires some prediction of the signal phase from the radio altimeter, as well as knowledge of the sweep frequency—however this is standardized. Since Mode 2 alerts are based on the rate of descent, the attacker can at least calculate the target change in round-trip time (RTT) to trigger an alarm. For example, descending at 3000 ft/min (≈ 15.4 m/s) at 500 ft AGL (≈ 152.4 m) will trigger an alarm according to standard (Fig. A2b in [51]). Using a simple model of the aircraft moving a negligible amount during a pulse, we use the difference in RTT over the course of one second (i.e. the aircraft at 152.4 m AGL, then one second later having descended 15.4 m). Eq. 1 then gives us the required change in is RTT, in which t_{rtt} is RTT, h is height above ground and c is the speed of light. This indicates a small jump in frequency per round-trip is needed.

$$\Delta t_{rtt} = \frac{2(h_1 - h_2)}{c} = \frac{2(152.4 - 137.0)}{c} \approx 1.03^{-7}\text{s/m} \quad (1)$$



(a) Protocol diagram for TCAS interrogation using the Mode S data link, where nearby aircraft respond with information on their position.



(b) Protocol diagram of TCAS all-call interrogation using Mode C, and response from nearby aircraft with altitude if available. Range and bearing are calculated from response.

Fig. 12: Representation of TCAS interrogation protocols of nearby aircraft using Mode C and S transponders.

An attacker will need a number of directional antennae underneath the approach path to transmit to the radio altimeter. These will be fed by SDRs; such equipment would be in the low $1000s. Although an attacker could operate such a system remotely, the hardware would need to be located near to the runway. The ability to deploy consequently depends on the airfield security and perimeter size, an analysis of which is not in our scope.

B. TCAS

TCAS makes use of the Mode C or Mode S transponders fitted to an object aircraft to interrogate nearby aircraft [17]. Establishing nearby aircraft with Mode S requires the object aircraft to listen for IDs in Mode S 'squitters', which are messages in response to ground-based Secondary Surveillance Radar (SSR) interrogations. The object aircraft can then interrogate these IDs to calculate whether nearby aircraft will become too close [18]. An abstracted protocol diagram can be seen in Fig. 12a.

Mode C operates differently, shown in Fig. 12b. The object aircraft issues Mode C-only interrogations called *all-calls*, causing all nearby aircraft with Mode C transponders to respond once a second with their altitude. Since Mode C does not carry the same data fields as Mode S, the object aircraft estimates range and bearing [20]. To limit interference, it uses a *whisper-shout* transmission mechanism, gradually increasing power and suppressing aircraft who already responded.

Attacks on TCAS are feasible as all discussed underlying transmissions are sent in the clear with no authentication. Schäfer et. al. exploit the weakness of such transponder transmissions [43] and the same fundamental wireless attack concepts can be translated to TCAS.

We firstly presume that we can establish the altitude, heading and speed of the target aircraft from broadcast surveillance messages [47]. The type of injection then depends on whether they use Mode S or Mode C:

- **Mode S**: the attacker transmits a false aircraft squitter message. When the target aircraft then interrogates, the attacker transmits Mode S responses as if the false aircraft were traveling on a collision course with the target.
- **Mode C**: the attacker responds to an all-call and following interrogations for the false aircraft. Whisper-shout may cause interrogations to be too low power to be received by the attacker, in which case they would need to approximate a response. However, this would be stochastic as interrogation rates are standardized.

The attacker can choose whether to cause the target to climb or descend by injecting an aircraft below or above the object aircraft respectively.

Transmission by the attacker would require an off-the-shelf amplifier and antenna capable of directional transmission, with a high-powered setup costing $15,000. A transceiver is needed to both receive interrogations to establish the target aircraft behavior, as well as transmitting false aircraft messages.

C. ILS/GS

No significant technical barriers exist for an ILS attack. This is possible due to the simplistic nature of the system—whilst it is monitored for integrity as defined in ICAO Annex 10, this is for deviations in the legitimate signal rather than malicious interference [27]. An ILS system will normally shut down or notify ATC if excessive deviation is identified.

An attacker will need an SDR, amplifier and directional antennae to replicate the antennae arrays used for the legitimate GS, costing around $10,000. Since no open-source tools exist to do this, software would need to be created but this is achievable by moderately resourced attacker as it involves implementing a standardized, static system. Furthermore, the transmission power is readily achievable with consumer amplifiers as a typical GS is below 10 W. For reference, even the lowest level of licensed UK amateur radio operators can transmit in frequency bands surrounding aviation bands at up to 10 W [35].

Related work suggests that ILS course deviation attacks are possible using such equipment. Sathaye et al. describe two signal generation approaches which enable ILS signal interference, leading to an attacker being able to adjust the localizer or glideslope as seen in the cockpit [40]. This is fundamentally the same type of attack as we test in the simulator, made possible with the same equipment and resource level as in our threat model.

Similar to our GPWS attack, the attacker will have to locate close to the airport perimeter to have correct signal directionality along the runway.

Appendix B
Debrief Interview Questions

Below are questions used in the debrief interview with participants. Each section was asked after participants had

flown that scenario, with the final section being asked at the end of the session. Scales for answers match to scale points provided in App. C, Tab. VI, e.g. 'Significant impact' matches to a score of 1.

GLIDESLOPE

1) During this scenario, did the aircraft perform as expected? [Yes/No]
   In particular, did the ILS approach happen as you would normally expect it to? [Yes/No]
   Briefly describe the impact of the ILS procedure not occurring as expected, particularly with respect to how this impacted flight and the steps you had to take to account for this.
   - I Significant impact
   - II Some impact
   - III Little impact
   - IV No impact
2) Did you opt to not use it? [Yes/No]
   - I If so would you use it later? [Yes/No]
3) How confident are you that this was the best decision in the circumstances?
   - I Very confident
   - II Somewhat confident
   - III Not sure
   - IV Not confident
   - V Significantly not confident
4) Do you feel that it put the aircraft in a less safe situation? [Yes/No]
   - I If so, how?
5) To what extent did this increase your workload?
   - I Significant increase
   - II Some increase
   - III No increase
6) Did this affect your trust in your systems?
   - I Much more trust
   - II Some more trust
   - III No effect
   - IV Some distrust
   - V Much distrust
7) If this happened in a real aircraft, do you feel you would act in the same way? [Yes/No]
   - I If not, which different steps would you take?

TCAS

8) During this scenario, did the aircraft perform as expected? [Yes/No]
   In particular, did the TCAS system behave as you would normally expect it to? [Yes/No]
   Briefly describe the impact of TCAS not behaving as expected, particularly with respect to how this impacted flight and the steps you had to take to account for this.
   - I Significant impact
   - II Some impact
   - III Little impact
   - IV No impact
9) Did you turn it off?
   - I If so would you turn it back on later? [Yes/No]
10) How confident are you that this was the best decision in the circumstances?
   - I Very confident
   - II Somewhat confident
   - III Not sure
   - IV Not confident
   - V Significantly not confident
11) Do you feel that it put the aircraft in a less safe situation? [Yes/No]
   - I If so, how?
12) To what extent did this increase your workload?
   - I Significant increase
   - II Some increase
   - III No increase
13) Did this affect your trust in your systems?
   - I Much more trust
   - II Some more trust
   - III No effect
   - IV Some distrust
   - V Much distrust
14) If this happened in a real aircraft, do you feel you would act in the same way? [Yes/No]
   - I If not, which different steps would you take? [Yes/No]

GPWS

15) During this scenario, did the aircraft perform as expected? [Yes/No]
   In particular, did the GPWS system behave as you would normally expect it to? [Yes/No]
   Briefly describe the impact of the GPWS not behaving as expected, particularly with respect to how this impacted flight and the steps you had to take to account for this.
   - I Significant impact
   - II Some impact
   - III Little impact
   - IV No impact
16) Did you turn it off? [Yes/No]
   - I If so would you turn it back on later? [Yes/No]
17) How confident are you that this was the best decision in the circumstances?
   - I Very confident
   - II Somewhat confident
   - III Not sure
   - IV Not confident
   - V Significantly not confident
18) Do you feel that it put the aircraft in a less safe situation? [Yes/No]
   - I If so, how?
19) To what extent did this increase your workload?
   - I Significant increase
   - II Some increase
   - III No increase
20) Did this affect your trust in your systems?

TABLE VI: Summary of participant interview responses for attack scenarios. Scale points are normalized so that 1 represents the most 'positive' point, i.e. the greatest change, and the highest value represents the most 'negative' i.e. no change. For example, using Q4 relating to impact, 1 is the 'significant impact' response. Dash indicates where no scale value existed, and representative scale point is taken as scale response at the rounded mean, e.g. for impact, 1.4 will be 'significant impact'.

		Number of Participant Responses per Scale Point											
Attack	Question	1	1.5	2	2.5	3	3.5	4	4.5	5	Mean	Representative Scale Point	Std. Dev
GS	Impact	10	3	10	2	4	0	1	-	-	1.85	Some impact	0.787
	Confidence	21	1	8	0	0	0	0	0	0	1.28	Very confident	0.441
	Workload	6	1	22	1	0	-	-	-	-	1.80	Some increase	0.420
	Trust	5	0	18	0	7	0	0	0	0	2.07	Some distrust	0.629
TCAS	Impact	19	3	5	2	1	0	0	-	-	1.38	Significant impact	0.573
	Confidence	12	4	11	0	3	0	0	0	0	1.63	Somewhat confident	0.639
	Workload	16	4	9	1	0	-	-	-	-	1.42	Significant increase	0.484
	Trust	19	2	8	0	1	0	0	0	0	1.36	Much distrust	0.531
GPWS	Impact	8	2	13	0	3	2	2	-	-	2.03	Some impact	0.894
	Confidence	24	0	5	0	1	0	0	0	0	1.23	Very confident	0.496
	Workload	13	2	11	3	1	-	-	-	-	1.62	Some increase	0.601
	Trust	10	3	16	0	1	0	0	0	0	1.65	Some distrust	0.519

I Much more trust
II Some more trust
III No effect
IV Some distrust
V Much distrust

21) If this happened in a real aircraft, do you feel you would act in the same way? [Yes/No]
   I If not, which different steps would you take?

FINAL DEBRIEF

22) Did you find the scenarios to be a useful exercise? [Yes/No]
23) To what extent did you feel limited by the simulator?
   a) Heavily limited
   b) Somewhat limited
   c) Not limited
24) Have you encountered any of the scenarios in the wild? [Yes/No] If so, provide detail.
25) Do you feel that this could be a useful training tool for pilots? [Yes/No]

APPENDIX C
INTERVIEW SCALE RESPONSE DATA

The full data for Fig. 7 can be seen in Tab. VI.