A Comprehensive Approach for Validation of Air Traffic Management Security Prototypes

A Case Study

Tim H. Stelkens-Kobsch, M. Finke, N. Carstengerdes
Institute of Flight Guidance
German Aerospace Center (DLR)
Braunschweig, Germany
Tim.Stelkens-Kobsch@dlr.de, Michael.Finke@dlr.de, Nils.Carstengerdes@dlr.de

Abstract— **Security in air traffic management is still a rather new challenge and receives increased interest during recent years. This implies that new security concepts and systems are developed. Usually all systems have to go through several validation cycles to reach a higher technical readiness level. As no well-established validation approach is available which considers the special aspects of security this forms an additional barrier when developing air traffic control security systems. This is true because suitable validation approaches have to be developed first. The latter includes the risk of forgetting something, when the development is not initiated in a structured way.**

Within the air traffic security project GAMMA such an approach has been developed and applied to a set of seven prototypes. Based on the European Operational Concept Validation Methodology and a Security Risk Assessment Methodology, this approach identifies additional security controls, system requirements, validation objectives and key performance indicators. These are the driving elements for the design of the validation setup and procedure

The paper demonstrates the feasibility of this new approach using one specific example, the Secure Air Traffic Control Communications prototype.

The paper describes the approach and the resulting validation setup and procedures in detail. It briefly describes the obtained results for the developed prototype as one specific use case of the approach.

Keywords-Air Traffic Management; ATM security; validation; ATC voice communication

I. Introduction

Safety research and implementation of appropriate measures to ensure a safe flow of air traffic is well established throughout the air traffic management (ATM) for quite some time [1]. One might remember the long way necessary to establish the indispensable safety management system procedures in ATM (hazard identification, risk management, performance measuring, safety assurance …). From the security point of view, comparable security management standards do not yet exist. Thanks to endeavors of recent years, the gap between highly sophisticated safety related and security related ATM research, which is still in its infancy, could be narrowed in the future. There are several scientific and commercial projects and initiatives intended to increase security in ATM. [2] [3]. One of the research projects to pave the way to enhance ATM security is the Global ATM Security Management Project (GAMMA, http://www.gamma-project.eu/) funded under the 7th Framework Program of the European Commission. GAMMA takes input from as well the Single European Sky ATM Research Program (SESAR) as Next Generation Air Transportation System (NextGen) and is intended to bring theoretical ideas developed in recent years down to practical implementations.

Within the project seven different prototypes for enhancing ATM security were developed. One of them, called Security Management Platform (SMP), can be seen as the core element of the GAMMA security management concept [4] and was developed to collect, correlate and disseminate security information within nations, from nations to European level and vice versa. The other six prototypes reside more on the system level and are intended for directly securing defined areas of interest within ATM. The different prototypes are intended to be used e.g. in internet applications adopted by air traffic management, integrated modular radios, satellite communications, Data link communications, Aeronautical Mobile Airport Communication Systems (AeroMACS) and Air Traffic Control (ATC) voice communications.

The structure of this document reflects the strategy applied to successfully conduct validations of ATM security prototypes. This structure can be understood as a blueprint for future validations of single ATM security prototypes and of prototype systems. After this short introduction, section II initially describes the context from which the work origins. Section III then explains the approach of the Security Risk Assessment Methodology (SecRAM) [5], which was used for identifying assets, vulnerabilities and threats. Section IV describes the purpose and design of a Secure ATC Communications (SACom) prototype as a technical example for an additional security control, whereas section V discusses the application of the SecRAM for this focus area in ATM. Section VI reports on the setup for the validation of the dedicated prototype. The validation is based on the well-known

The research leading to the results presented in this paper has received funding from the European Union FP7 under grant agreement n° 312382. More information can be found under http://www.gamma-project.eu/.

European Operational Concept Validation Methodology (E-OCVM) [6]. Section VII discusses the results and finally section VIII gives some conclusions and a short outlook regarding the proposed methodology for validating ATM security systems.

II. Context and Scope

A. *Context*

One feasible approach to describe the operational context when detailing the work conducted is to look from a management point of view. Management services in air transportation are categorized according to ICAO into Air Traffic Management (ATM), Communication, Navigation and Surveillance (CNS), Meteorological Services (MET), Aeronautical Information Services (AIS) and Search and Rescue (SAR). Fig. 1 depicts this context in an illustrative way (SAR is left out for simplification) [7].

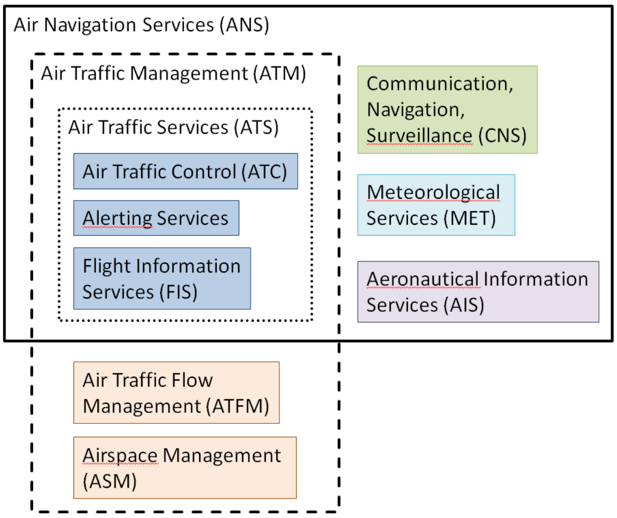

Air Navigation Services (ANS)
Air Traffic Management (ATM)
Air Traffic Services (ATS)
Air Traffic Control (ATC)
Alerting Services
Flight Information Services (FIS)
Communication, Navigation, Surveillance (CNS)
Meteorological Services (MET)
Aeronautical Information Services (AIS)
Air Traffic Flow Management (ATFM)
Airspace Management (ASM)

Figure 1. Components of ATM [7]

CNS/ATM is seen as the core service nucleus (or system) for the provision of air traffic management services (i.e. airspace management (ASM), air traffic flow management (ATFM), and air traffic services (ATS)) [8].

MET, AIS and SAR are considered as external to ATM (MET, AIS and SAR organizations are responsible for the security of their systems and functions themselves).

Interfaces to MET, AIS and SAR organizations and interoperability with associated systems fall within the scope of ATM Security.

In order to facilitate the understanding of the context the classification of different security topics will be carried out.

B. *Scope*

Fig. 2 shows a possible distinction between different focus areas of security. Fig. 2 has to be understood as a qualitative statement; the overlap areas are neither true to scale nor claiming completeness.

Aviation Security may be subdivided into ATM Security, Aircraft Security and Airport Security. The research presented herein will focus on the ATM Security.

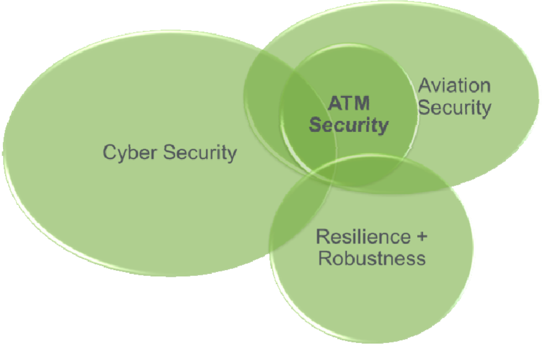


Figure 2. Relation of selected security areas

III. Risk Assessment and Treatment Methodology

In order to describe the primary assets residing in the frame of ATM systems and being affected by attacks a thorough investigation has to be undertaken (see also [9]).

The procedural steps needed are guided by several methodologies. Following SecRAM (Fig. 3), two types of primary assets have to be taken into account: (1) services and (2) (primary) information.

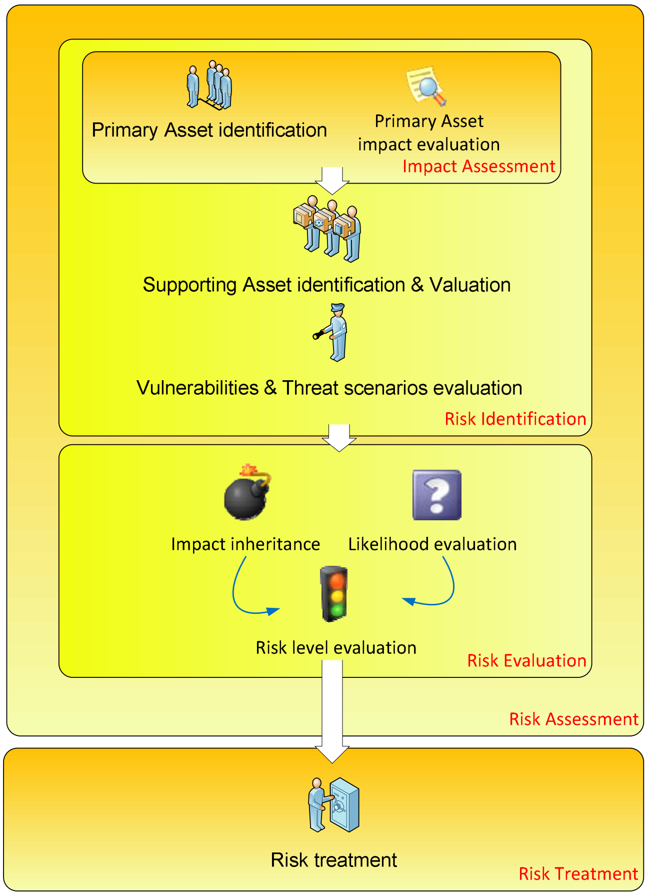


Figure 3. SecRAM process overview [5] [10]

(1) The services can be sub-divided in (a) services addressed by Operational Focus Areas (OFA), (b) system services, (c) operational concepts and operational activities and (d) necessary services to comply with contractual, legal or regulatory requirements.

(2) Information is considered as primary, when it is (i) vital for exercise of mission or business, (ii) personal regarding privacy, (iii) strategic or confidential and (iv) high-cost (regarding duration of acquisition or plain cost).

After the primary assets have been identified the possible impact on the level of Confidentiality, Integrity and Availability (CIA) has to be assessed (Table I). The impact has to be evaluated regarding both loss and degradation of the asset under investigation. In order to evaluate the consequences the security impact areas defined by SecRAM need to be used.

Hereafter the supporting assets need to be named. Supporting assets are tangible elements within the scope that support the existence of the primary assets. Typically these elements are e.g. entities involved in storing, processing and/or transmitting primary assets. Examples are servers, databases, laptops and workstations. In Air Traffic Control the voice communication can be seen as one of its supporting assets. The relation of supporting assets to primary assets may be described as 1-N (each supporting asset is linked with one or more primary assets).

TABLE I. IMPACT EVALUATION

Supporting asset	**Impact evaluation**									
	Threat	***Primary asset N***			***Primary asset N+1***			…	***Impact***	
		C	***I***	***A***	***C***	***I***	***A***	…	***Inherited***	***Reviewed***
SA #01	Threat #01	x		x		x		…	5	4
SA #01	Threat #02		x	x			x	…	3	2
…	…	…	…	…	…	…	…	…	…	…

TABLE II. LIKELIHOOD REGISTRATION

Supporting asset	**Likelihood registration**				
	Threat scenario	***Reviewed impact***	***Exposure level***	***Potentiality level***	***Rationale for parameters***
SA #01	Threat #01	5	4	3	This kind of scenario would trigger regional/national media attention. Expertise and knowledge required makes it somehow likely to occur.
SA #01	Threat #02	…	…	…	…
…	…	…	…	…	…

After identification of the supporting assets it is needed to reveal the vulnerabilities which could be exploited by adversaries. This step in the process inherits deep expert knowledge in order to identify the weak points in the ATM system. Now that vulnerabilities have been found the associated threats which endanger the system confidentiality, integrity and availability need to be conceived. Then the related risks that the prevailing vulnerabilities can be exploited have to be assessed. This is achieved by using the presented guidance material and obtaining expert knowledge. Thereafter the likelihood that the supporting assets are affected by a threat needs to be rated. This is done by using Table II and Table III.

TABLE III. LIKELIHOOD EVALUATION

Supporting asset	**Likelihood evaluation**		
	Threat	***Reviewed impact***	***Likelihood***
SA #01	Threat #01	5	3
	Threat #02	5	2
SA #02	…	…	…
…	…	…	…

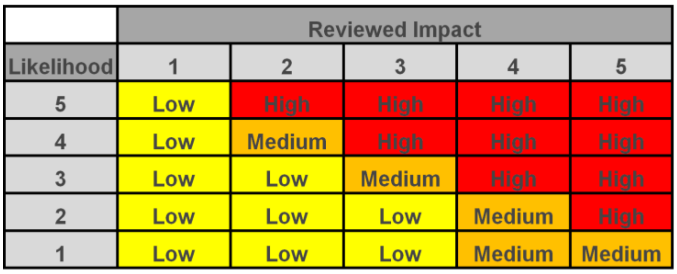
With Table III the likelihood that a threat occurs is considered and rated. Table IV is then intended to estimate the risk level (low, medium, high) by taking into account the previous considerations. Presented here is a snippet of the table, which was established within the GAMMA project.

TABLE IV. RISK LEVEL EVALUATION

Supporting asset	**Risk level evaluation**			
	Threat	***Reviewed impact***	***Likelihood***	***Risk level***
SA #01	Threat #01	5	3	High
	Threat #02	5	2	High
SA #02	…	…	…	…
…	…	…	…	…

The resulting risk values are now composed in a matrix, which is called risk level evaluation (Fig. 4).

This risk matrix can be used to define needed measures for reducing the risk appetite and scaling down the likelihood that a threat is successful (risk treatment). Furthermore the impact of a successful threat can be decreased. The pending steps are now to postulate security objectives to secure the assets. The security objectives represent the measures chosen when working with the risk matrix. This means for each asset of concern one or several security objectives are identified. The security controls shall ensure that remaining residual risks still existing after the treatment meet the postulated security objectives for the assets. For the GAMMA project this is documented in [11].



	Reviewed Impact				
Likelihood	1	2	3	4	5
5	Low	High	High	High	High
4	Low	Medium	High	High	High
3	Low	Low	Medium	High	High
2	Low	Low	Low	Medium	High
1	Low	Low	Low	Medium	Medium

Figure 4. SecRAM risk matrix [5]

The risk treatment needs to be supported by well-known catalogues of generic descriptions of security controls. In aviation a preferred suggestion is to use e.g. the Minimum Set of Security Controls (MSSC) developed by SESAR [12]. After the risk minimization effect of these controls has been rated, there might be a residual risk. This residuum now is treated by additional security controls (ASC), which are systems, assets or procedures not yet existing. At this point the need for postulating new security prototypes and/or procedures arises. Within the research project GAMMA for example this evaluation led to the development of seven different ATM security prototypes developed by different partners of the project. The detailed approach chosen for the application of the SecRAM methodology in GAMMA is shown in Fig. 5 and documented in [13]. The numbers represent the quantity of items identified.

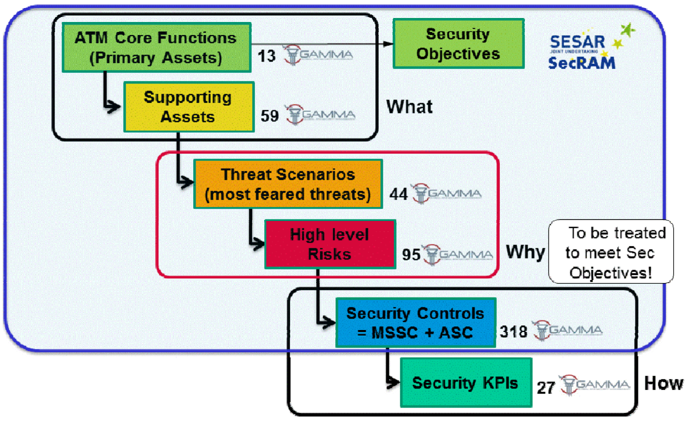

ATM Core Functions (Primary Assets)
13
Security Objectives
SESAR SecRAM
Supporting Assets
59
What
Threat Scenarios (most feared threats)
44
High level Risks
95
Why
To be treated to meet Sec Objectives!
Security Controls = MSSC + ASC
318
Security KPIs
27
How
GAMMA

Figure 5. Security Risk Assessment and Treatment in GAMMA [14]

The prototypes mainly focus on separate subdivisions in ATM. Six of the prototypes reside on system level and one prototype is intended to collect, rate and disseminate the information received by the others. The latter is called SMP, (briefly described in section I). The six system level prototypes are (1) Information Exchange Gateway (IEG), (2) SATCOM Security, (3) Information Security System (ISS), (4) Secure GNSS Communication, (5) Integrated Modular Communications (IMC) and (6) Secure ATC Communications (SACom) [15].

In the remainder of this paper the application will be demonstrated by taking the SACom prototype as a specific example. The SecRAM application provides the basis for the intended development of a security prototype validation methodology.

In order to set the scene the reasons for the development of such a prototype will be explained.

IV. The Need for Secure ATC Communications

The rationale for developing a security prototype in the area of ATC voice communications is underpinned by the fact that radio communication used by ATC can easily be intruded and has therefore been subject to recurrent attacks [16] [17] [18]. Nevertheless the voice communication between pilots and air traffic controllers is still the basic and most important communication method within the aeronautical mobile service, as it is the most flexible and efficient medium especially in a busy traffic environment or when non-standard situations occur.

The communication via voice is therefore one of the supporting assets of ATC. Although the use of data link is steadily increasing, it can only partly replace air-ground voice communication and cannot be used in time-critical and/or non-standard situations with current datalink procedures. For example, CPDLC is not designed for aerodrome control, approach control or VFR flights and does not provide sufficient response times [19].

The international voice communication standards in aviation involve the use of omnidirectional analogue radio transceivers with double-sideband and amplitude modulated carrier waves. The VHF frequency band range is defined with 117.975-137.000 MHz [20].

The commonly known characteristics of analogue omnidirectional radio voice communication resulting from wave propagation physics can be summarized as following:

- Requires line-of-sight to a certain extent (e.g. a ground-based transmitter may be received by an airborne radio but not by another ground-based receiver; the same airborne transmitter is most likely received at both ground stations (Fig. 6)).
- Simultaneous transmissions on the same frequency cause interference making both transmissions (almost) unreadable (so called "block-out").
- The reception quality decreases with increasing distance to the transmitter (nearby stations may outgo stations of a larger distance).
- Due to the analogue nature, modern encryption technology cannot be used.

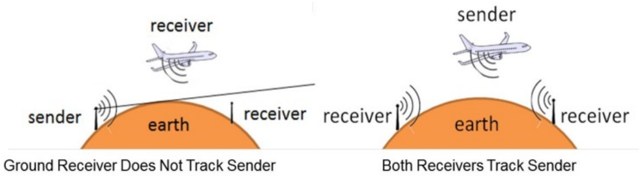

receiver
sender
earth
receiver
Ground Receiver Does Not Track Sender
sender
receiver
earth
receiver
Both Receivers Track Sender

Figure 6. ATC Voice Communications [21]

As a conclusion, the still widely used analogue air-ground radio voice introduces a significant security risk:

- Physically, the current unsecured air-ground voice communication can be freely accessed.
- Appropriate radio communication equipment to access the air-ground voice communication is available for purchase with almost no restriction.
- Any unlawful interference, especially those that appear to be credible transmissions, may remain undetected for a certain time.
- Any interference directly influences the provision of Air Traffic Control Service and therefore also directly endangers the assurance of preventing collisions between two aircraft or between an aircraft and another object and of maintaining a safe, expeditious and orderly flow of air traffic.

Possible security threats targeting the air-ground voice communication (see Table I – IV and [10]) are:

- Intentional frequency blocking/jamming,
- Fake transmissions to airplanes by unauthorized persons (in the following referred to as "False ATCO", i.e. False Air Traffic Controller) with the goal to severely disturb the safe and orderly flow of air traffic. Such fake transmissions are not necessarily received by Air Traffic Control due to the effects described above.

The SACom prototype addresses the second one of the above mentioned security threats. Recent examples for this kind of unlawful interference were reported in [16] and [18] and impacts were investigated in more detail in [16].

This prototype consists of several modules providing the following functions:

- Verification of all speakers in air-ground voice communication by analyzing voice characteristics ("speaker verification function").
- Determination of the current stress level by analyzing stress-typical voice characteristics in air-ground voice communication ("stress detection function").
- Trajectory-based and state-vector-based conformance monitoring to detect aircraft deviating from given ATC clearances ("conformance monitoring function").
- Trajectory-based and state-vector-based conflict detection ("conflict detection function").
- Correlation of the functions above as a basis for automatic reports to security management instances.

V. Applying SecRAM for SACom

The above described methodology has been applied and the primary and supporting assets have been identified for the validation of Secure ATC Communications. The primary assets which need to be protected (or at least parts of) were found as:

- Communication service.
- Arrival management and landing procedure.
- Pre-departure sequencing, departure management and take-off procedure.
- Conflict management (separation/collision avoidance).

After the primary assets have been determined the following step was to identify the supporting assets. The result of the determination is shown below:

- Voice system.
- En route Area Control Center (ACC).
- Each approach control unit.
- Each aerodrome control tower.

Subsequently the threats able to exploit the vulnerabilities of the supporting assets needed to be imagined. This appears to be the most critical part in the risk assessment, which also requires a high level of expert knowledge about the functionality of the supporting assets.

For these threats a detailed investigation regarding impact evaluation, likelihood registration, likelihood evaluation and risk level evaluation has been conducted following SecRAM [5] and described in section III, which provides the described set of tables to fulfill this task.

In the first risk treatment step the security controls already in place need to be named by identifying them from a catalogue of pre-described controls. Such a catalogue is, for instance, inherited in the MSSC.

Taking the mitigation effects of those more general security controls into account there may be not enough success to reduce the risk level to an acceptable value. However, this can be further enhanced by proposing adapted general security controls (taken from the MSSC) or additional security controls which need to be invented from scratch. When implementing these controls it is assumed that the risk level is mitigated to acceptable values (i.e. low or medium). The needed ASC to achieve the mitigation are listed in Table V.

The list of additional security control drives the postulation of requirements for the validation setup. The prototype can now be developed and verified according to these constraints. Hereafter the validation exercises can be planned.

TABLE V. Security control list SACom

Supporting asset	**Security control list**		
	Threat	***Security control ID***	***Control description***
Voice system	False ATCO	ASC_TFA_01	Air-Ground voice system shall change to digital radio communication which provides means to encrypt messages
	Freq. blocking	ASC_TFB_01	
Voice system	False ATCO	ASC_TFA_02	Air-Ground voice system needs a restricted access to hardware
	Freq. blocking	ASC_TFB_02	
Voice system	False ATCO	ASC_TFA_03	Air-Ground voice system shall use message encryption
Voice system	False ATCO	ASC_TFA_04	Each ACC/TWR shall provide means to detect unusual trajectory of flight
Voice system	False ATCO	ASC_TFA_05	Air-Ground voice system shall be supported by means to detect voice pattern anomaly
Voice system	False ATCO	ASC_TFA_06	Each ACC/TWR shall operate and control speaker verification
Voice system	Freq. blocking	ASC_TFB_06	Air-Ground voice system shall have the possibility to increase transmitting power of transmitter

VI. Validation Setup

Within the GAMMA project the SESAR guidance material available for risk assessment and treatment has been applied straightforward. The consecutive step was then to apply also the E-OCVM to plan and execute tailored validations in a structured way. The combination of these methodologies delivers the inevitable blueprint for the validation of ATM security prototypes.

According to the GAMMA Validation Exercise Plan [22] (which was written using the guidance of E-OCVM) the functionalities of the SACom prototype are proven by validating the single prototype in standalone configuration. Following the E-OCVM, system requirements and, in the following, the validation objectives were derived.

In order to develop distinct security validation scenarios, the baseline was defined as the existing operational concepts and system functionalities with respect to security. Against this baseline, the GAMMA benefits were demonstrated and validated.

In order to validate any requirement, Key Performance Indicators (KPI) need to be defined which provide a measurement of efficiency to weigh e.g. the benefit of an additional security control and to assess if security objectives (e.g. performance of the prototype, acceptance, enhancing of situational awareness) are fulfilled. Several indicators and values were determined during the validation trials (see also [23]):

- Performance of the prototype's speaker verification function (Detection Rate/False Alarm Rate).
- Performance of the prototype's stress detection function (Detection Rate/False Alarm Rate).
- Performance of the prototype's conformance monitoring function (Detection Rate/False Alarm Rate) as well as the air traffic controllers performance practicing the monitoring function without any support (Detection Rate).
- Performance of the prototype's conflict detection function (Detection Rate/False Alarm Rate) as well as the air traffic controllers performance practicing the conflict detection function without any support (Detection Rate).
- User Acceptance.

The definitions of detection rate and false alarm rate in this context are depending on the application area of each prototype module and will be detailed in the results and discussions part

The chosen validation method for the SACom prototype was Human-in-the-loop (HITL) simulations. Within the study six air traffic controllers participated with more than ten years of controllers' experience, four of them were male. Five of the participants were mid-aged experienced German ATC center controllers, whereas one controller was a mid-aged person and experienced as well in Australian as Irish ATC centers. Each person under test acted as an ATCO and was confronted with many different events, caused by security and/or safety problems. As a prerequisite to conduct the validation exercises voice examples of all persons under test needed to be recorded (a so called speaker enrollment).

The validation exercise duration was 8 hours, spread over two days. The exercise started with a briefing, introducing the research topic and also explaining the goals of the general security concept. However, the participants were unaware about the False ATCO threat and the validation questions of the following exercise runs. Afterwards, the ATCOs had to be enrolled and authorized in a speaker database for the SACom prototype. Then the participants had about 20 minutes to familiarize with the simulation and their controller working position. During the remainder of day one, the participants acted as approach controllers in 20 short scenarios with duration of about five minutes each. In some of these scenarios different threats occurred and the ATCOs had to cope with them. In the background and unnoticed from the participants the SACom prototype was running and thus creating both baseline data together with performance data from the participants. On the second day the SACom prototype was explained to the participants without mentioning that in the following exercise scenarios a False ATCO attack will be performed. Hereafter a training scenario was performed where the participants could test all functionalities. Afterwards there was one simulation run of 45 minutes duration, where the SACom indications and warnings were visible to the participants and False ATCO attacks were performed. Afterwards a long debriefing was conducted which included several questionnaires.

As a summary, one complete exercise consists of the following steps:

1) Briefing of the participants
2) Speaker verification enrollment and enrollment test
3) Simulator training (no SACom indications visible)
4) 20 short simulation scenarios (no SACom indications visible)
5) SACom Briefing
6) SACom training simulation (SACom indications visible)
7) False ATCO attack simulation (SACom indications visible)
8) Debriefing and questionnaires

VII. Results and Discussion

A. Performance of the Speaker Verification Function

The speaker verification function of the SACom prototype delivers a score value (ranging from 0 to 100) for each transmission of any speaker. Authorized speakers usually showed speaker verification scores of 40-70 while unauthorized speakers usually showed scores of less than 30.

Validation exercise step 2) was used to determine the optimum alert threshold for unauthorized speakers (transmissions with a measured speaker verification score values at and below the alert threshold are considered as unauthorized).

For evaluation the following definitions have been taken:

- The detection rate was defined as percentage of all unauthorized transmissions which were detected as unauthorized transmission.
- The false alarm rate was defined as percentage of all authorized transmissions which were wrongly classified as unauthorized transmission.

The results shown in the Table IV were obtained from approximately 100 utterances spoken per exercise run involving all speakers of the exercise (three authorized and one unauthorized speaker).

The values in Table VI have been gained under the following conditions:

- All speakers had no secondary tasks during this exercise step and had the opportunity to fully concentrate on giving the utterances.
- No time constraint was present.
- A limited number of utterances were considered.
- The used Voice over IP (VoIP) audio system provided a very high audio quality.

TABLE VI. SPEAKER VERIFICATION RESULTS

Exercise Run	**Aircraft conformance – Detection Rate**		
	Optimum Alert Threshold	***Detection Rate***	***False Alarm Rate***
1	15	100%	3.3%
2	14-30	100%	0%
3	27-35	100%	0%
4	21	96.0%	0%
5	31-40	100%	0%
6	33	91.7%	0%

Recapitulating the results presented above, a very high reliability is shown (very high detection rate of about 91.7% to 100% and very low false alarm rate of about 0% to 3.3%). Provided that the system is robust against the named factors, the potential to apply the speaker verification algorithms of the SACom to air-ground voice communication is clearly visible. This is true especially to directly detect unauthorized transmissions in the frame of the above mentioned threat scenario.

It has been observed that the time difference between the end of the transmission and the display of the result plays a critical role for the usability of this function because a human operator must always be able to correlate the audio transmissions with the indications. If the time difference between both events is too large a human operator will not be able to identify which utterance caused an alert.

During the exercises, a default alert threshold setting of 30 was used as a first estimation. After completing the run, the alert threshold were be adjusted to achieve the best results. This optimum alert threshold shows differences from exercise run to exercise run. Hence, an alternate solution needs to be found to (continuously) adapt depending on the actual constellation of speakers and used audio equipment.

Another constraint is the speaker enrollment. This voice example should have a length of 3-5 min of continuous speech and should be recorded with the audio equipment that is going to be used in the radio conversation. Factors like fatigue, stress or sickness as well as a reduced audio quality had significant influence on the performance of the speaker verification function in the exercises. For implementing this function in the existing air-ground voice communication a solution has to be found for managing the speaker enrollments for a very large number of authorized participants and either the audio quality of radio communication or the robustness of the used speaker verification function has to be improved.

B. Stress Detection Function

The stress detection function of the SACom prototype delivers a stress score for each transmission of any speaker. The stress score is determined by searching for stress-typical voice patterns according to a database of stressed speech. The value of the stress score should directly reflect the experienced level of stress. Validation exercise steps 3) and 7) were used to obtain stress scores within simulations were it is expected that the speakers are relaxed (to determine a false alarm rate) and within simulations were it is expected that the speakers are under stress in some predefined situations (to determine a detection rate).

The detection rate was defined as the percentage of all transmissions which are assumed to be under stress and which are correctly classified as stressed transmissions.

The false alarm rate was defined as the percentage of all transmissions which are assumed to be free of stress and which were wrongly classified as stressed transmissions.

At first it has to be mentioned that in the frame of the project it was not possible to check the success of the stress induction by using well-established means like psychophysiological measures or questionnaires, as the resulting effort and budget was not covered.. This fact has to be kept in mind when interpreting the results.

Usually the determined detection rates are below 30% with a large variation between the different exercises. The determined false alarm rates show the same trend (false alarm rates of up to 30% with large variation between the different exercises. The reliability of stress detection based on voice pattern analysis as implemented with the SACom prototype may directly depend on the voice characteristics themselves; the reliability may then differ from person to person.

According to the statements above, further research is necessary to raise the stress detection to a mature state. One of the most important steps would be the creation of a database of stressed and unstressed controller speech samples to extract typical voice stress patterns in ATC (controllers and pilots). Such a database is far away from being available.

C. Conformance Monitoring

The conformance monitoring function of the SACom prototype delivers an indication for each aircraft when the system detects a deviation from the given clearance. This function needs precise, correct, complete and actual information about given clearances, which can be gathered e.g. from highly sophisticated speech recognition tools. This would avoid incorrect, missing or late clearance inputs which jeopardize correct functioning.

Validation exercise step 4) was used to obtain the performance of the prototype and of the air traffic controller monitoring the traffic without any assistance. This was done by carefully reviewing simulation recordings and correlating the deviations detected by SACom with the traffic situation.

1) Detection Rates

The detection rate (for both SACom and ATCO performance) was defined as the percentage of existing aircraft deviations which were correctly detected by the prototype resp. the air traffic controller.

Table VII shows obtained results for the detection rates.

TABLE VII. CONFORMANCE MONITORING DETECTION RATE

Partici-pant	Aircraft conformance – Detection Rate		
	ATCO Detection Rate	*SACom Detection Rate*	*Difference SACom - ATCO*
1	92.0%	88.0%	-4%
2	76.7%	93.3%	+16.6%
3	91.7%	95.8%	+4.1%
4	80.0%	96.7%	+16.7%
5	84.6%	88.5%	+3.9%
6	85.0%	85.0%	0%

The distribution shows that the system performance is basically equal or higher than the performance of the air traffic controller. This shows the potential of conformance monitoring assistance tools in general both from the security and the safety point of view.

Simulated events which the controller frequently did not notice were:

- Level deviations (because - in contrast to lateral deviations which are directly visible on the radar screen - the controller has to read and process the altitude information displayed in the radar label).
- Deviations which do not directly conflict with the preplanning of the air traffic controller.
- Deviations of aircraft which do not yet need any guidance (in the used approach control simulation e.g. shortly after handover still far away from the airport).

One short simulation run contained a planned level deviation during the final leg of the ILS approach, induced by a simulated false advice from an unauthorized speaker to discontinue approach, climb on runway heading flight level 70. All six controllers did not detect the deviation in time, as a level deviation was unexpected after the aircraft reported to be established on the final approach. The situation ended in 2 near-miss situations and 4 mid-air collisions with another approaching aircraft (Airborne collision avoidance systems were not simulated). This underlines the severity of such a security threat.

2) False Alarm Rate

One feasible possibility to determine a false alarm rate is to calculate the number of false alarms divided by the number of all alarms. In doing so a false alarm is defined as an alert by the prototype’s conformance monitoring system without any deviation of the considered aircraft from the spoken ATC clearance. This approach has been chosen for evaluation of the conformance monitoring to be the most appropriate.

The false alarms can be categorized by carefully reviewing simulation recordings and pseudo pilot command logs. Two types of false alarm rates can be defined:

- False Alarm Rate Type 1: Number of false alarms including false alarms caused by incorrect, missing or late clearance inputs divided by the number of all alarms,
- False Alarm Rate Type 2: Number of false alarms excluding false alarms caused by incorrect, missing or late clearance inputs divided by the number of all alarms.

Table VIII shows the results for the false alarm rates. It is obvious that the system is very vulnerable against wrong, missing or late clearance inputs, causing a high number of false alarms (e.g. participant 2 had a false alarm rate type 1 of 71%). After these errors have been eliminated, it can be seen that roughly 10% of all alarms are false alarms (mainly caused by simulation errors), which would be acceptable.

This underlines the need for a very reliable method to precisely and quickly capture the spoken ATC clearance. In this validation exercise, speech recognition technology which is available at the German Aerospace Center (DLR) research facility in Braunschweig was used. This speech recognition technology showed very high recognition rates in former validation trials [24] [25]. According to the results presented here, this can be considered as the absolute minimum required performance for using speech recognition technology together with monitoring tools in general. This underpins the need to use the all modules of SACom in combination to achieve a system which can increase the security level in ATM.

TABLE VIII. CONFORMANCE MONITORING FALSE ALARM RATE

Partici-pant	Aircraft conformance – False Alarm Rate	
	False Alarm Rate Type 1	*False Alarm Rate Type 2*
1	63.8%	10.3%
2	71.0%	7.2%
3	34.2%	2.9%
4	56.7%	9.0%
5	58.9%	12.5%
6	52.8%	11.1%

3) Average Time until Detection

The average time until detection was determined for the SACom prototype as well as for the Air Traffic Controller without any technical support. For every single aircraft deviation, the time difference between the simulation timestamp at which the deviation was visible for the first time on the radar display and the simulation timestamp at which the system or the air traffic controller recognized the deviation was determined. Finally, the average was calculated over all single deviations for every exercise run, which is summarized in Table IX.

These results clearly show that an automatic monitoring system – provided that it is fed with complete, reliable and valid clearance information – is able to detect aircraft deviations much faster than an air traffic controller. In these experiments, the system was always between 20 and 30 seconds faster than the unsupported air traffic controller. There were several situations during the simulations where it would have made a significant difference in safety to detect the event 20 seconds earlier. This shows again the potential of conformance monitoring tools in general.

TABLE IX. CONFORMANCE MONITORING TIME UNTIL DETECTION

Partici-pant	Aircraft conformance – Time until Detection		
	ATCO Average Time until Detection	*SACom Average Time until Detection*	*Difference SACom - ATCO*
1	41.6 s	16.5 s	-25.1 s
2	39.4 s	11.8 s	-27.6 s
3	43.1 s	15.8 s	-27.3 s
4	38.7 s	14.5 s	-24.2 s
5	38.9 s	13.9 s	-25.0 s
6	34.7 s	14.1 s	-20.6 s

D. *Conflict Detection Task*

The conflict detection function of the SACom prototype delivers an indication for each aircraft constellation where the system detects a risk of a loss of separation. Besides the "classical" conflict detection according to the actual speed vectors, the SACom prototype also uses context information to predict aircraft trajectories. Similar to the conformance monitoring function this needs precise, correct and complete information about given clearances.

Validation exercise step 4) was used to obtain the performance of the prototype and of the air traffic controller searching for possible conflicts without any assistance. This was done by carefully reviewing simulation recordings and correlating the conflicts detected by SACom with the traffic situation.

A conflict was defined as a situation where the usual minimum IFR separation in approach sectors (3NM lateral separation and 1000ft vertical separation) is not ensured within the next 60 seconds.

The detection rate (for both SACom and ATCO performance) was defined as the percentage of existing conflicts which were correctly detected by the prototype resp. the air traffic controller.

The false alarm rate was defined as the number of SACom conflict alerts without any real conflict situation as defined above (taking the latest ATC clearances into account) divided by the number of all SACom conflict alerts.

During the simulation runs, the conflict detection function of the SACom prototype showed a similar performance for the detection rate as the air traffic controller. The average detection rate over all exercise runs for both was about 85%.

The false alarm rate for the SACom prototype showed values between 0% and 20%.

E. *User Acceptance*

The air traffic controllers taking part in these exercise runs gave their feedback and were asked to fill out both bespoke and well-established, standardized validation questionnaires. These questionnaires covered not just the rating of SACom but also about the simulation setup, realism, the GAMMA concept and also questions about ATM security in general.

Exemplary, Fig. 7 shows an extract of the results obtained for the user acceptance. The blue pillars show the mean ratings while the black brackets show the standard deviation of the answers.

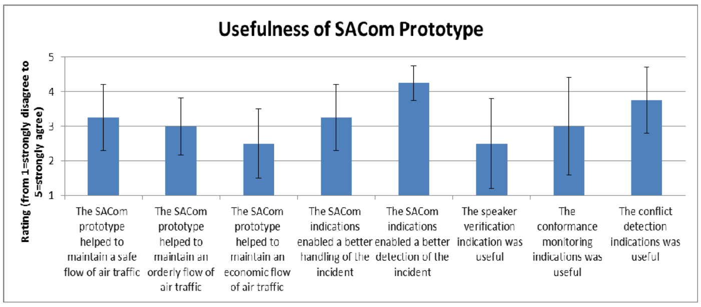


Figure 7. User ratings obtained from questionnaires

The obtained data show a general agreement to the approach taken with the SACom prototype. A closer inspection of the available answers reveals that there is a clear trend in the data concerning a distinguished view of the various modules of the SACom prototype. The Conflict Detection module always gets the best assessment, followed by the Conformance Monitoring module. The Speaker Verification Module was rated as the least useful feature, mainly due to the chosen method of presentation and not regarding the function as such (speaker verification results were presented in an additional window and showing every verification result and not only unrecognized speaker alarms). This feedback will be taken into account for the further development of the HMI of the prototype. It has to be noted that the neutral opinions of the participants regarding statements concerning the safe and orderly flow of traffic and the slightly negative remarks concerning an economic flow of traffic cannot be seen as a negative result. Rather this result was expected as the prototype was meant to enhance security without the intent to improve the safe and orderly flow of traffic. According to the participants the economic flow of traffic suffered a little bit, however, this comes as no surprise and was expected as a security trade-off. In summary, the participants agreed that the SACom prototype enabled a better detection of the False ATCO attack and is a preferable solution to secure ATC voice communication.

No questions were asked concerning the stress detection feature, because stress detection results were not displayed to the controllers during the exercise.

Summarizing the results of the extensive debriefing sessions, the participants had a positive view of the SACom prototype and its modules, seeing benefits of the prototype itself and the GAMMA concept in general for improving ATM security. The concept was accepted and areas of improvement for some modules were identified.

VIII. CONCLUSIONS

The aim of this paper was to describe a methodology to build and validate ATM security prototypes. This was implemented by combining well-known methodologies like SecRAM and E-OCVM. The result can be used as a blueprint for successful security prototype validation. This approach was exemplified using a dedicated prototype. The conclusions can be divided into prototype-specific results and the practicability of the elaborated methodological approach.

Regarding the prototype results it has to be kept in mind that for SACom it is very hard to clearly separate and distinguish security events from safety events based on software algorithms solely. Aircraft (hence pilots) deviating from a given ATC clearance may do this because of safety reasons (e.g. loss of control) and/or security reasons (e.g. hijacking). Detection systems like the SACom can hardly distinguish between both with only one indicator (e.g. aircraft deviations). Only a correlation of several indicators can identify a security incident [26]. As a fundamental finding a system like SACom will be useful for safety purposes, too. During validation, the SACom prototype clearly showed potential as an assistance system for handling the simulated events, especially the conformance monitoring function and the enhanced conflict detection function. Both functions need continuous, correct, complete and reliable updates about given ATC clearances, which underlines again the potential of the combination of such tools with speech recognition.

Concerning the practicability of the elaborated methodology, this approach seems to be straightforward and promising. The achieved results foster the idea to postulate a comprehensive methodology for validating ATM security systems and ATM security prototypes. Both SecRAM and E-OCVM methodologies provided practical assistance for setting up the validations. Not only the needed security control could be elaborated but also a relevant validation exercise was established.

Following the facts and methodological steps a blueprint for validation of ATM security prototypes looks as follows:

- Identify the problem, PA, SA, threats and vulnerabilities.
- Gather PA, SA and analyze risk by applying SecRAM.
- Identify relevant KPI and prototype requirements.
- Build up the prototype.
- Postulate validation objectives.
- Invent scenarios for validation.
- Evaluate the prototype according to E-OCVM.

The comprehensive approach for validation of air traffic management security prototypes has been conducted the first time within the ATM security research by the project GAMMA.

Acknowledgment

The authors would like to thank all GAMMA consortium members that contributed to this paper through stimulating discussions around the concepts presented.

References

[1] ICAO, “Safety Management,” Annex 19 to the Convention on International Civil Aviation, 1st Edition, 2013.

[2] CSFI ATC Cyber Security Project, www.csfi.us, July 2015.

[3] EU_ECAC CASE project, https://www.ecac-ceac.org/ec-ecac-case-project.

[4] GAMMA Consortium, 2015, GAMMA CONOPS, Rev. 01.00, http://www.gamma-project.eu/wp-content/uploads/2013/11/GAMMA-Concept-of-Operations_Rev-01-00.pdf.

[5] SESAR Joint Undertaking, "SESAR ATM Security Risk Assessment Methodology," - Project 16.02.03 D02, 2013.

[6] EUROCONTROL, 2010, European Operational Concept Validation Methodology, Version 3.0, https://www.eurocontrol.int/publications /european-operational-concept-validation-methodology-eocvm.

[7] Kreuz, M., "Modellierung von Flugsicherungsprozessen auf Basis von System Dynamics," Forschungsbericht/DLR, Deutsches Zentrum für Luft- und Raumfahrt, 2015, 33.

[8] Kölle, R., Proceedings International Summer School on Aviation Psychology (ISAP), Graz July 2007.

[9] P. Montefusco, R. Casar, R. Koelle, T. H. Stelkens-Kobsch, “Addressing security in the ATM environment: from identification to validation of security countermeasures with introduction of new security capabilities in the ATM system context,” ARES, 2016 11th, 532-541.

[10] GAMMA consortium, 2015, D2.1 – Threat analysis & evaluation report.

[11] GAMMA consortium, 2015, D2.2 – Security objective report.

[12] Minimum Set of Security Controls, SESAR Project 16.02.05, D05-006, Edition 00.06.00, August 2013.

[13] GAMMA consortium, 2015, D2.3 – Risk treatment report.

[14] P. Montefusco, “GAMMA Security Risk Assessment and Treatment,” presentation for advisory board of GAMMA, Brussels, December 2015.

[15] T. H. Stelkens-Kobsch, A. Hasselberg, T. Mühlhausen, N. Carstengerdes, M. Finke, C. Neeteson, „Towards a more secure ATC voice communications system,“ Digital Avionics Systems Conference (DASC), 2015 IEEE/AIAA 34th, DOI: 10.1109/DASC.2015.7311419.

[16] M. Strohmeier, M. Schaefer, R. Pinheitro, V. Lenders, I. Martinovic, “On Perception and Reality in Wireless Air Traffic Communications Security,” arXiv preprint arXiv:1602.08777, DOI: 10.1109/TITS.2016.2612584.

[17] LiveATC, “Fake ATC in Action (LTBA – Istanbul)”, May 2011, http://www.liveatc.net/forums/atcaviation-audio-clips/25-may-2011-fake-atc-in-action-%28ltba-istanbul%29.

[18] The Age, “Lone-wolf radio hoaxer hacks Melbourne air traffic control”, November 2016, http://www.theage.com.au/ victoria/lonewolf-radio-hoaxer-hacks-melbourne-air-traffic-control-afp-20161107-gsk12o.html.

[19] ICAO, “Procedures for Air Navigation Services - Air Traffic Management,” Doc 4444, 15th Edition, 2007.

[20] ICAO, “Communication Systems,” Annex 10 to the Convention on International Civil Aviation Vol. III, 2nd Edition, 2001.

[21] C. Neeteson, M. Rusko, “WP6 Secure ATC Communication (SACom),” presentation at GAMMA WP6 kick off meeting, Rome, February 2015.

[22] GAMMA consortium, 2015, D5.1 – Validation exercise plan.

[23] T. H. Stelkens-Kobsch, M. Finke, D. Kolev, R. Koelle, R. Lahaije, „Towards validating a security situation management capability,“ Integrated Communications Navigation and Surveillance (ICNS), 2016, 1A1-1-1A1-9, DOI: 10.1109/ICNSURV.2016.7486320.

[24] Helmke, H., Rataj, J., Mühlhausen, T., Ohneiser, O., H. Ehr, H., Kleinert, M., Oualil, Y, Schulder, M. and Klakow, D., “Assistant-based speech recognition for ATM applications,” 11th FAA/EUROCONTROL ATM-seminar, Lissabon, Portugal, June 2015.

[25] H. Helmke, O. Ohneiser, T. Mühlhausen, M. Wies, "Reducing Controller Workload with Automatic Speech Recognition", 35th Digital Avionics System Conference, Sacramento, CA, USA, September 2016.

[26] T. H. Stelkens-Kobsch, M. Finke, M Kleinert, M. Schaper, „Validating an ATM security prototype – first results,“ DASC, 2016 IEEE/AIAA 35th, DOI: 10.1109/DASC.2016.7778107.