Frequency-based detection of replay attacks: application to a quadrotor UAV

Conference Paper - October 2019
DOI: 10.1109/ICSC47195.2019.8950619

4 authors, including:

Damiano Rotondo
Spanish National Research Council
103 PUBLICATIONS  867 CITATIONS

Joseba Quevedo
Universitat Politècnica de Catalunya
233 PUBLICATIONS  2,531 CITATIONS

Some of the authors of this publication are also working on these related projects:

Fault diagnosis and fault tolerant control for uncertain nonlinear systems View project

Fault diagnosis of nonlinear Takagi-Sugeno systems View project

All content following this page was uploaded by Damiano Rotondo on 24 October 2019.

The user has requested enhancement of the downloaded file.
Frequency-based detection of replay attacks: application to a quadrotor UAV

Helem S. Sánchez, Damiano Rotondo, Marc López Vidal, Joseba Quevedo

Abstract—Unmanned aerial vehicles (UAVs) are reported to be highly exposed as possible targets of cyber attacks, due to their strong strategic and economic value, and their increasing use in a wide range of operations. Among the most critical cyber attacks, replay attacks are performed by replacing the real data coming from the sensors with previously recorded data, causing deterioration of the control system’s performance and potentially allowing other types of attacks without being discovered. The main contribution of this paper is to investigate the applicability of a frequency-based detection method, which uses a sinusoidal signal with a time-varying frequency as authentication signal, to a UAV affected by replay attacks. The effectiveness of the method is illustrated through simulation scenarios.

I. INTRODUCTION

In recent years, the increasing integration of cyber (computation, communication, etc.) and physical processes has led to the introduction of a new class of systems, known as cyber-physical systems (CPSs) [1]. CPSs are characterized by a higher efficiency, but also by bigger vulnerabilities, which can be exploited by a malicious agent in order to perform a cyber attacks, resulting in some kind of damage or economical loss [2], [3].

Unmanned aerial vehicles (UAVs) are reported to be highly exposed as possible targets of cyber attacks, due to their strong strategic and economic value, and their increasing use in a wide range of operations, such as border surveillance, reconnaissance, transport, and even civilian tasks, as e.g. fun by hobby enthusiasts [4]. Among the attacks reported for UAVs, we recall the one performed in 2012 during a flight test in South Korea, in which a GPS jamming attack was executed on a S-100 Camcopter, resulting in a crash into the ground control van, which lead to the killing of an engineer and the injuring of two remote pilots [5].

Due to the importance of assessing the vulnerabilities of UAVs, and developing new defense mechanisms, which could help in mitigating them, simulations of cyber attacks on UAVs are being executed [6]. [7] performed a GPS spoofing attack on a quadrotor, resulting in the UAV landing at an incorrect location. A deauthentication attack on a quadrotor was used by [8] to demonstrate its vulnerability, resulting in the loss of control by the operator. The work [9] showed that a hacker can exploit shared data and predictable collision avoidance properties in order to control and alter the UAV’s trajectory.

Among the most critical cyber attacks, there are the replay attacks. When an attack of this type is carried out, at first the attacker records the measurements coming from the sensors. Later, the attacker replaces the real data with the recorded one, causing deterioration of the control system’s performance and potentially allowing other types of attacks without being discovered. The threats to the security of UAVs brought by this type of attack have been well documented, see e.g. [10], [11]. In the last few years, different approaches have been proposed to detect these attacks, which may be classified roughly into two categories: i) watermarking-based approaches [12]–[16], in which an authentication signal is added to the control signal (at the cost of sacrificing the control performance), and the received sensor measurements are analyzed to check whether there is or there is not the effect of the authentication signal on the physical system; and ii) alternative methods [17], [18], which try to detect replay attacks without injecting signals in the control input.

The work [16] presented a method to detect replay attacks using a frequency-based signature, which was applied to a multiple tank system. The main contribution of the present paper is to investigate its applicability to a UAV affected by replay attacks. In particular, the method introduces a sinusoidal signal with a time-varying frequency (authentication signal) into the closed-loop system, and checks whether the time profile of the frequency components in the output signals are compatible with the authentication signal or not, by comparing the energies of appropriate signals.

The remaining of the paper is structured as follows. In Section II, the frequency-based replay attack detection technique is presented. Section III describes the application of the proposed method to a quadrotor. Section IV presents the simulation results. Finally, Section V provides the conclusions.

II. REPLAY ATTACK DETECTION METHOD

A. Attack definition and overview of the detection method

Replay attacks corrupt the measurements coming from the sensors, and are carried out in two stages [12]:

1) in the first stage, the attacker records sensor data $y(t)$ in a time window $[t_0, t_0 + w]$, where $w$ denotes the attack duration, without compromising their integrity;
2) In the second stage, the recorded data are used to replace the real data, such that the controlled system is exposed to some kind of harm without the supervisory unit/agent being aware of it, e.g., the system could be driven to a different operating point, or some act of theft/sabotage could be performed.

Hereafter, we recall the method proposed in [16], which detects replay attacks (also identifying which output channels are being attacked) by introducing a sinusoidal signal with a time-varying frequency into the system. Then, the detector makes a decision by checking if the measured output is compatible with the introduced signature or not. More specifically, by applying a dynamical decoupling technique based on vector fitting [19], it can be ensured that a given input channel will affect only an output channel, which allows identifying which channel is being attacked. Finally, the detector compares the energies of band-pass signals to determine if a replay attack is being carried out or not.

B. Signal generation

Let us consider a continuous-time linear time invariant system:

\[ \dot{x}(t) = Ax(t) + Bu(t) \]  
\[ y(t) = Cx(t) \]

where \( x \in \mathbb{R}^n \) is the state, \( u \in \mathbb{R}^m \) is the input, \( y \in \mathbb{R}^p \) is the output, and \( A, B, C \) denote matrices of appropriate dimensions. The system (1)-(2) is controlled by means of a state-feedback control law:

\[ u(t) = -Kx(t) \]

such that the closed-loop system is described by:

\[ \dot{x}(t) = (A - BK)x(t) \]

The frequency-based method for replay attack detection introduces a signature \( \zeta(t) \) into the input \( u(t) \), such that:

\[ u(t) = -Kx(t) + \zeta(t) \]

It is straightforward to see that \( \zeta(t) \) will affect the output \( y(t) \) according to the transfer function:

\[ G(s) = C(sI - A + BK)^{-1}B \]

which usually contains coupling, i.e., a signature signal \( \zeta_i(t) \) introduced in the \( i \)-th input channel will contribute to the response of all the output signals contained in \( y(t) \). Since we would like to use the signature signals to identify which channel is affected by the replay attack, we introduce a decoupler \( F(s) \) with input \( \tilde{\zeta}_i(t) \) and output \( \zeta_i(t) \), such that the series interconnection \( G_d(s) = G(s)F(s) \) is a decoupled system, i.e., made up by a diagonal transfer matrix. Then, we can introduce frequency-varying sinusoidal signals:

\[ \zeta_i = \alpha_i \cos(\omega_{\sigma_i(t)}t) \quad i = 1, \ldots, m \]

into the decoupler, where \( \alpha_i \) denotes the signals’ magnitudes, and \( \sigma_i(t) \) denote piecewise constant signals, taking integer values between 1 and \( N \), such that at each instant of time \( \omega_{\sigma_i(t)} \) take values within a set of predetermined frequencies \{\( \omega_1, \ldots, \omega_N \}\}, according to the value of \( \zeta_i(t) \). The idea behind introducing (7) is that by analyzing the output signals, we can decide about the attack occurrence by checking if the received output is compatible with the introduced signals or not.

It is clear that, since a complete decoupling of \( G(s)F(s) \) is very demanding, and in some cases cannot be achieved at all [20], we can require decoupling for the finite set of frequencies \( \omega_i, i = 1, \ldots, N \). More specifically, by requiring that \( F(\omega_i) = G(\omega_i)^{-1} \), a set of \( N \) constraints that the decoupler should satisfy is obtained. Then, the robust numerical method for rational approximation known as vector fitting can be applied [21].

C. Detection logic

The output signal \( y(t) \) will contain the effects of the different inputs to the process, e.g., the input signal \( u(t) \). With the aim of analyzing only the content due to the signature signal \( \tilde{\zeta}_i(t) \), \( y(t) \) is processed through a bank of band-pass filters (each filter corresponding to a different frequency \( \omega_i, i = 1, \ldots, N \)), whose transfer functions are obtained as follows [22]:

\[ H_i(s) = \text{diag} \left\{ \frac{\omega_i}{s + \omega_i} \right\} \]

where \( Q_i \) is the selectivity of the filter, bigger values of which correspond to a narrower frequency response, but also a slower dynamic response.

Let us denote as \( z_i(t) \) the output of the filter \( H_i(s) \) fed by \( y_l(t) \). Then, we can obtain an estimate \( \tilde{\sigma}_i(t) \) of \( \sigma_i(t) \) by comparing the different signals \( z_i(t) \). As proposed in [16], a possible way to obtain \( \tilde{\sigma}_i(t) \) is to compare the energies of \( z_i(t) \) over the largest period associated with the frequencies \( \omega_i \):

\[ T_{\omega_i} = \max_{i=1,\ldots,N} \frac{2\pi}{\omega_i} \]

while also taking into account the time \( t_{\text{trans}} \) needed after a switch in \( \tilde{\sigma}_i(t) \) in order for the transient to become negligible. Then, the signals \( \tilde{\sigma}_i(t) \) are calculated as (10) (see top of the next page), where \( t_{\text{trans}}^* = \frac{1}{T_{\omega_i}} T_{\text{trans}} \) denotes the last switching time. Then, the replay attack detection algorithm will compare the known signals \( \tilde{\sigma}_i(t) \) with the estimated signals \( \tilde{\sigma}_i(t) \). If a mismatch is found, the algorithm will warn about \( y_l(t) \) being corrupted by a replay attack.

III. APPLICATION TO A QUADROTOR

In this section, we present the application of the frequency-based replay attack detection method to a quadrotor.

A. Linearized model

The quadrotor is a vehicle that has four rotors in a cross configuration. Two rotors can rotate in a clockwise direction, while the other two can rotate anticlockwise. The four rotors are controlled independently, and their speeds affect the overall movement of the quadrotor. It is assumed that the quadrotor has a rigid and symmetrical structure, with
\[
\hat{\sigma}(t) = \begin{cases} 
\sigma(t) \\
\arg \max_{\tau = t} \int_{t}^{t+T_a} |z_{d}(\tau)|^2 d\tau 
\end{cases} \quad \text{if } \sigma(t) \neq \sigma(t-T_a) \wedge t \in [t_s^r, t_s^a + t_{trans} + T_a] \]
\[
\text{otherwise}
\]

center of gravity that coincides with the body fixed frame, and rigid propellers, which provide thrust and drag forces that are proportional to the square of the propellers’ speeds.

Let us introduce the state vector \( x = [z, \dot{z}, \phi, \dot{\phi}, \theta, \dot{\theta}, \psi, \dot{\psi}]^T \) and the input vector \( u = [\Omega_1, \Omega_2, \Omega_3, \Omega_4]^T \), where \( z \) is the height, \( \phi, \theta, \psi \) are the roll, pitch and yaw angle, respectively, and \( \Omega_1, \Omega_2, \Omega_3, \Omega_4 \) are the rotor speeds. By applying a constant feedforward input \( u_c = [\Omega_1, e, \Omega_2, e, \Omega_3, e, \Omega_4, e]^T \), with:

\[
\Omega_1, e = \Omega_2, e = \Omega_3, e = \Omega_4, e = \sqrt{\frac{gm}{4b}}
\]

a hovering equilibrium point is obtained, characterized by a constant equilibrium state \( x_e \), in particular, the equilibrium height is a value \( z_e \), and the equilibrium angles \( \phi_e, \theta_e \) and \( \psi_e \) are all zero. Then, by considering deviations \( \Delta x \) of the state variables from \( x_e \), and performing a linearization of the nonlinear state equations (see [23], [24]), a model akin to (1) is obtained:

\[
\Delta \dot{x}(t) = A \Delta x(t) + B \Delta u(t)
\]

where:

\[
A = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
0 & -\frac{I}{\sqrt{g m}} & 0 & 0 & 0 & 0 & 0 & 0 \\
\sqrt{g m} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\frac{I}{\sqrt{g m}} & 0 & 0 & 0 & 0 & 0 & 0 \\
\sqrt{g m} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -\frac{I}{t^2} \sqrt{g m} & 0 & 0 & 0 & 0 & 0 \\
\sqrt{g m} & 0 & 0 & -\frac{I}{t^2} \sqrt{g m} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\frac{I}{t^2} \sqrt{g m} & 0 & 0 & 0 & 0 \\
\sqrt{g m} & 0 & 0 & 0 & -\frac{I}{t^2} \sqrt{g m} & 0 & 0 & 0 \\
\end{bmatrix}
\]

which, together with a linear error-feedback control law \( \Delta u(t) = -K \Delta x(t) \), leads to the closed-loop system:

\[
\Delta \dot{x}(t) = (A - BK) \Delta x(t)
\]

We will assume that both the altitude and the attitude of the quadrotor are monitored by a ground station. The communication between the UAV and the station can be hacked through a replay attack. Hence, the system’s description (1)-(2) is completed by the output matrix \( C \), which is given by:

\[
C = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\end{bmatrix}
\]

TABLE I

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l_1 )</td>
<td>8.1e-3</td>
<td>Nms^2</td>
</tr>
<tr>
<td>( l_2 )</td>
<td>8.1e-3</td>
<td>Nms^2</td>
</tr>
<tr>
<td>( l_3 )</td>
<td>14.2e-3</td>
<td>Nms^2</td>
</tr>
<tr>
<td>( m )</td>
<td>1</td>
<td>kg</td>
</tr>
<tr>
<td>( g )</td>
<td>9.81</td>
<td>ms^-2</td>
</tr>
<tr>
<td>( J )</td>
<td>104e-6</td>
<td>Nms^2</td>
</tr>
<tr>
<td>( b )</td>
<td>54.2e-6</td>
<td>Ns^2</td>
</tr>
</tbody>
</table>

or, equivalently, by defining \( \Delta y(t) = y(t) - [z_e, \phi_e, \theta_e, \psi_e]^T \):

\[
\Delta y(t) = C \Delta x(t)
\]

Also, for the remaining of the manuscript, we will assume that the quadrotor’s parameters are the ones provided in Table I, where \( l_1, l_2, l_3 \) denote the body inertia along \( x, y, z \) axes, \( m \) is the quadrotor’s mass, \( g \) denotes the gravity acceleration, \( J \) is the propeller inertia, \( l \) is the lever’s length, and \( b \) and \( d \) are the thrust and drag factor, respectively.

B. Choice of the design parameters

Let us consider a replay attack detector, as described in the previous section, with \( N = 2 \) and \( \omega_2 = 2\omega_1 \). Due to the fact that by design:

\[
G(\omega) \approx F(\omega)^{-1} = \left[ C(\omega - A + BK)^{-1} B \right]^{-1}
\]

the following relationship holds between the magnitudes \( \alpha_1, \alpha_2, \alpha_3, \alpha_4 \) of \( \tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3, \tilde{\xi}_4 \) (input signals to the decoupler) and the maximum magnitudes \( \alpha_1, \alpha_2, \alpha_3, \alpha_4 \) of \( \xi_1, \xi_2, \xi_3, \xi_4 \) (outputs of the decoupler, which are inputs of the actuators);

\[
\left[ \alpha_1 \alpha_2 \alpha_3 \alpha_4 \right] = \max_{i=1,2} \left[ C(\omega - A + BK)^{-1} B \right]^{-1} \left[ \alpha_1 \alpha_2 \alpha_3 \alpha_4 \right]
\]

where the max is understood element-wise. By requiring that:

\[
\alpha_i \leq \kappa_i u_{e,l}, l = 1, \ldots, n_u
\]

where \( \kappa_i \ll 1 \) and \( u_{e,l} \) is the \( l \)-th element of \( u_e \), such that the increment of the input signal brought by the signature signal is relatively small, a set of feasible frequencies is calculated, and \( \omega_1 \) is chosen as the maximum value among these frequencies. The values \( \omega_i, i = 1, \ldots, 4 \) are design parameters whose values should be chosen small enough such that \( \tilde{\xi}(t) \) is not clearly visible in the output signals, and big enough to overcome the effect of exogenous disturbances and measurement noise. In the following, we select \( \kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = 1/10 \) and \( \omega_1 = 0.005 \) and \( \alpha_2 = \alpha_3 = \alpha_4 = 0.01 \).

Four different controllers \( K_1, K_2, K_3, K_4 \) have been considered in order to study how the particular choice of the controller affected both the set of feasible frequencies for
TABLE II
CLOSED-LOOP POLES POSITION

<table>
<thead>
<tr>
<th>Controller</th>
<th>Dominant poles</th>
<th>Fast poles</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_1$</td>
<td>$(-3,-3,-3,-3)$</td>
<td>$(-10,-10,-10,-10)$</td>
</tr>
<tr>
<td>$K_2$</td>
<td>$(-3,-3,-3,-3)$</td>
<td>$(-20,-20,-20,-20)$</td>
</tr>
<tr>
<td>$K_3$</td>
<td>$(-3,-3,-3,-3)$</td>
<td>$(-50,-50,-50,-50)$</td>
</tr>
<tr>
<td>$K_4$</td>
<td>$(-3,-3,-3,-3)$</td>
<td>$(-100,-100,-100,-100)$</td>
</tr>
</tbody>
</table>

$\omega_1$ and the decoupling performance of the decoupler $F(s)$. All the controllers have been designed by pole placement, by requiring that the closed-loop poles were placed in different positions of the complex plane, as resumed in Table II. Eq. (14) has been used to calculate the expected $\alpha$ (due to symmetry of the quadrotor, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$) for the different controllers, such that (15) can be checked graphically, as shown in Fig. 1. Using the controller $K_1$, a value $\omega_1 = 5.1 rad/s$ has been found, while with the controllers $K_2$ and $K_3$, $\omega_1 = 3.3 rad/s$ and $\omega_1 = 0.3 rad/s$ have been obtained. On the other hand, no feasible frequency $\omega_1$ was found for controller gain $K_4$. Although in principle it is desired to obtain a value of $\omega_1$ as high as possible, so that $T_0$ in (9) becomes smaller, and the detection algorithm based on estimating $\hat{\sigma}_i(t)$ using (10) behaves faster, the choice of $\omega_1$ should take into account the decoupler performance which, looking at (13), depends on the specific controller gain (the decoupler can be obtained applying the VFIT3 routine\(^1\) [19], [25], as described in [16]).

In particular, Fig. 2 shows the Bode plot of the transfer function from the input 1 to the output 1 of the cascade $G(s)F(s)$, for the different controller gains $K_1$, $K_2$, $K_3$. Note that the values on the x-axis are normalized with respect to $\omega_1$, such that $\omega_{\text{norm}} = 1 rad/s$ corresponds to $\omega = 5.1 rad/s$, $\omega = 3.3 rad/s$ and $\omega = 0.3 rad/s$, respectively. It can be seen that the best decoupling performance is achieved with the controller $K_3$. However, using $K_3$ would lead to $\omega_1 = 0.3 rad/s$ and, through (9), to $T_0 = 20.94 s$, ultimately leading to a very slow attack detector. For this reason, using $K_2$ ($\omega_1 = 3.3 rad/s$, $T_0 = 1.91 s$) is deemed to be a good tradeoff between a fast detector and a good decoupling, and will be the controller used in the next section. By requiring an attenuation of $-20 dB$ at frequency $\omega_2$ for the first band-pass filter, and at frequency $\omega_1$ for the second band-pass filter, the selectivity parameters are calculated as $Q_1 = Q_2 = 2\sqrt{\frac{T}{T_0}}$. Also, $t_{\text{trans}}$ is chosen as the biggest among the settling times of the band-pass filters $H_i(s)$, and $T_i$ is chosen as $T_i = 4t_{\text{trans}}$, as in [16], thus obtaining $t_{\text{trans}} = 15.87 s$ and $T_i = 63.48 s$. For the sake of completeness, Fig. 3 shows the comparison between the magnitude Bode plot of the non-decoupled (blue line) system and the decoupled one (red line). It can be seen that the series interconnection $G(s)F(s)$ approximates an identity matrix, such that a good decoupling is achieved.

IV. SIMULATION RESULTS

In order to assess the effectiveness of the proposed strategy, two simulation scenarios are considered. In Scenario 1, the attacker records the measurements of the first output during the first 200 s, and replays periodically the data starting from $t = 200 s$. The arising mismatch between the outputs of the band-pass filters $z_{1i}(t)$, $z_{2i}(t)$ and the time-varying frequency profile of $\omega_1\sigma_i(t)$ (see Fig. 4) allows detecting the attack after 8.21 s, as shown in Fig. 5, which depicts the logical assessment of $\sigma_i(t) = \sigma_i(t)$, $i = 1,2,3,4$. In Scenario 2, the attacker corrupts the measurements $y_2(t)$ and $y_3(t)$ instead. In this case, the mismatches $\sigma_2(t) \neq \sigma_2(t)$ and $\sigma_3(t) \neq \sigma_3(t)$ allow identifying correctly the output channels as being attacked at time 398.65 s and 208.21 s, respectively (see Figs. 6-7).

V. CONCLUSIONS

This work has investigated the applicability of a frequency-based replay attack detection method to a quadrotor UAV. In particular, the method introduces a sinusoidal signal with a time-varying frequency into the closed-loop
system, and checks whether the time profile of the frequency components in the output signals are compatible with the authentication signal or not, by comparing the energies of appropriate signals. The simulation scenarios have shown that the method can be applied to the considerate case study, being capable of not triggering false alarms and identifying correctly the attacked output channels.

REFERENCES

Fig. 4. Scenario 1. Outputs of the band-pass filters $z_{il}(t)$ and time-varying frequency $\omega(t)$.

Fig. 5. Scenario 1. Result of the replay attack detection test.

Fig. 6. Scenario 2. Outputs of the band-pass filters $z_{il}(t)$ and time-varying frequency $\omega(t)$.

Fig. 7. Scenario 2. Result of the replay attack detection test.