An Asymmetric Security Mechanism for Navigation Signals

Markus G. Kuhn

University of Cambridge, Computer Laboratory,
15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom
http://www.cl.cam.ac.uk/~mgk25/

Abstract. Existing navigation services, such as GPS, offer no signal-integrity (anti-spoof) protection for the general public, especially not with systems for remote attestation of location, where an attacker has easy access to the receiver antenna. With predictable broadcast signals, the antenna can be replaced with a signal generator that simulates a signal as it would be received elsewhere. With a symmetrically encrypted broadcast signal, anyone who can build or reverse engineer a receiver will know the secret key needed to spoof other receivers. Such encryption is only of use in closed user communities (e.g., military) or with highly tamper-resistant modules protecting the common key. In open user communities without common secret keys, integrity protection is needed instead, with properties similar to digital signatures. The ability to verify a navigation signal must be separate from the ability to generate a new one or to apply selective-delay attacks; but simply signing the broadcast signals will not protect their exact relative arrival times. This paper introduces a practical solution based on short-term information hiding.

1 Introduction

Alice runs a transport company for high-valued goods. Her armoured lorries are equipped with satellite navigation receivers. These are queried via radio every few minutes by her computer. If one of her lorries deviates from the planned route or loses contact without plausible explanation, she can take action immediately to prevent it being stolen.

Bob runs a prison service. Some of his "clients" live and work outside the prison, but have to remain within a specified area. Others are offenders on probation who must stay outside certain areas or just have their location monitored continuously. Bob attaches a navigation receiver to their ankles and his prison computer queries that via radio (e.g., GSM) several times per hour.

Several such systems for remote attestation of location via the Global Positioning System (GPS) have been fielded, in particular for vehicle tracking [1]. The use of trusted GPS receivers has also been proposed for location-based network authentication [2]. Radio tagging of offenders to control a curfew is now practised

J. Fridrich (Ed.): IH 2004, LNCS 3200, pp. 239–252, 2004.

in several countries [3].[1] Other potential applications include road-charging and tachograph systems.

These are examples of security systems that use a navigation-signal receiver as a trusted component. Such a receiver may end up in the hands of an attacker with a strong incentive to manipulate the system such that it reports a *pretended position* $\mathbf{r}'$ instead of its *actual position* $\mathbf{r}$.

Section 2 below very briefly reviews the operating principles of modern positioning systems, Sect. 3 describes different classes of attacks on trusted positioning receivers, and Sect. 4 reviews briefly the symmetric security mechanisms available to military users of GPS and a technique proposed by Denning and MacDoran [2]. Section 5 then presents a new information-hiding based defense against the selective-delay attack from Sect. 3. Unlike previously proposed techniques, it adds to navigation signals an asymmetric security property known from digital signatures, namely that those able to verify the integrity of an antenna signal are not able to synthesize one that could pass the same verification process. Sect. 6 discusses a variant of the selective-delay attack involving directional antennas and how to defend against it, and Sect. 7 finally illustrates how some of the parameters involved might be chosen in a practical implementation.

2 Conventional Pseudorange Positioning Systems

Modern positioning systems use a number of transmitters X_i located at known coordinates $\mathbf{x}_i \in \mathbb{R}^3$. Each transmitter is equipped with a synchronized clock and knows the exact system time t. A receiver R is located at the coordinates $\mathbf{r} \in \mathbb{R}^3$ (to be determined). If each transmitter X_i broadcasts a navigation signal $s_i(t)$ that propagates through space in all directions with speed c, then we will receive at position $\mathbf{r}$ the signal

$$g(\mathbf{r}, t) = \sum_i A_i \cdot s_i \left(t - \frac{|\mathbf{x}_i - \mathbf{r}|}{c} \right) + n(\mathbf{r}, t) \tag{1}$$

where A_i is the attenuation the signal suffers on its way from X_i to R, and $n(\mathbf{r}, t)$ is background noise (see Fig. 1). With carefully chosen functions $s_i(t)$ (low auto- and cross-correlation, include timestamps and information on transmitter position), the receiver can separate the individual terms of this sum, identify the time delay $|\mathbf{x}_i - \mathbf{r}|/c$ for each and infer from it the “range”

$$d_i = |\mathbf{x}_i - \mathbf{r}| \ . \tag{2}$$

With three known ranges d_i to known transmitter positions $\mathbf{x}_i$, three equations (2) can be solved unambiguously for $\mathbf{r}$ (unless all three $\mathbf{x}_i$ are located on a line).

[1] Due to the difficulties of receiving satellite signals indoors, most offender tagging systems still rely on a base station installed in the monitored person’s home. However, future global positioning systems with increased transmitter power, lower carrier frequencies and improved receiver technology (e.g., long integration times) may well work reliably enough indoors to be used in such applications.

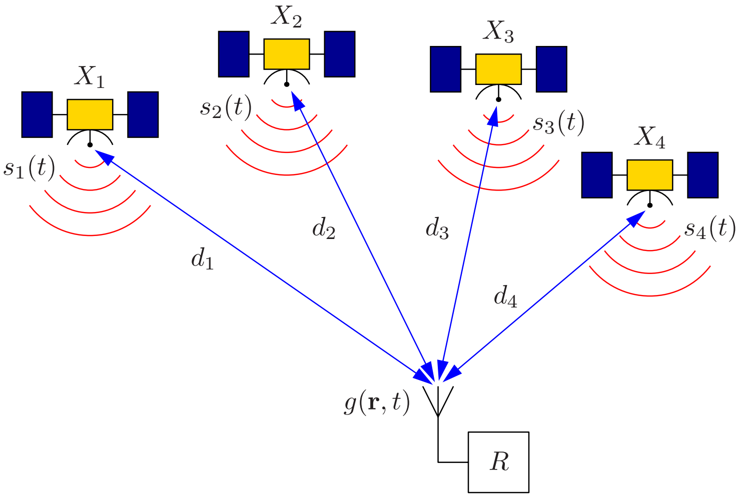

Fig. 1. A pseudorange navigation receiver R works by observing at its position $\mathbf{r}$ the delayed broadcast signals $s_i(t - d_i/c)$ from at least four transmitters X_i. Their relative delays can be used to solve four equations that determine the 3-dimensional position $\mathbf{r}$ and the time t.

Highly stable clocks (e.g., caesium oscillators) are costly and pure receivers cannot participate in two-way clock synchronization. Therefore, in practice, R will only have access to an imprecise estimate $t_R = t + u_R$ of the exact system time t. It therefore receives the signal

$$g(\mathbf{r}, t_R) = \sum_i A_i \cdot s_i \left(t - \frac{|\mathbf{x}_i - \mathbf{r}|}{c} + u_R \right) + n(\mathbf{r}, t_R) \tag{3}$$

and can infer from the delays $|\mathbf{x}_i - \mathbf{r}|/c - u_R$ only the “pseudoranges”

$$\tilde{d}_i = |\mathbf{x}_i - \mathbf{r}| - c \cdot u_R \ \ . \tag{4}$$

The clock error u_R adds a fourth unknown scalar. With pseudorange measurements to at least four transmitters X_i, the resulting system of equations (4) can be solved for both $\mathbf{r}$ and u_R, providing both the exact position and time, without requiring a precise local clock.

3 Attacks on Navigation Receivers

We now consider an attacker of a system for remote attestation of location who has access to its navigation receiver (for example, because it was tied to her ankle following a court order). There are two points to manipulate.

The first is the output of the receiver or the channel over which it reports the position of its antenna. The receiver could be substituted with a device that continuously outputs pretended positions $\mathbf{r}'$. This can be prevented with well-understood cryptographic authentication protocols that protect the link to the querying computer. If the receiver is only moderately tamper-resistant, an attacker who successfully extracts the key used in one will not have gained anything useful for spoofing the location reports from other receivers, making this attack difficult to scale. We are not concerned with such attacks in this paper.

The second point of attack is the navigation antenna, or more generally speaking, the connection of the receiver with the electromagnetic environment specific to its location. An attacker can separate the antenna from the receiver, or equivalently place it into a shielded enclosure along with a transmitting antenna, either way gaining full control over the input of the receiver. This enables several types of attack on a tamper-resistant receiver whose output is cryptographically protected.

In a *relaying attack* (also known as *worm-hole attack*), the receiver is connected to a remote antenna located at the pretended position $\mathbf{r}'$.[2] Such an attack may be logistically complex (arrangements may have to be made to move the remote antenna around in a plausible way) and the remote antenna can easily be located. One possible countermeasure might involve the use of a high-bandwidth signal, to maximize the cost of forwarding it. Another might use a highly stable clock in the receiver, to detect the signal delay introduced by a relaying attack. We are not concerned with relaying attacks in the rest of this paper.

In a *signal-synthesis attack*, the receiver is connected to a device that generates the navigation broadcast signal $g(\mathbf{r}', t)$ as it can be expected to be found at the pretended location. With fully-standardized plaintext broadcast signals, where all aspects of the message format and modulation are publicly known, a modest amount of hardware can simulate the signal to be expected at any point in time and space.

The obvious countermeasure against the signal-synthesis attack is to encrypt the individual broadcast signals $s_i(t)$, such that the attacker cannot predict the waveform $g(\mathbf{r}', t)$ that the receiver needs to see before it can report its position as $\mathbf{r}'$.

Carefully implemented encryption can guarantee the integrity and confidentiality of transmitted data, but this alone is not sufficient in the case of a navigation signal. Here the security-critical aspect of the signals $s_i(t)$ lies not only in the data they carry, but also in their exact relative arrival times at the receiver.

This is exploited in the *selective-delay attack*, in which the attacker uses the signal $g(\mathbf{r}, t)$ received at the actual position $\mathbf{r}$, converts it into a prediction of the signal $g(\mathbf{r}', t - \Delta t)$ that would have been received at the pretended position $\mathbf{r}'$ a short time Δt earlier, and feeds that into the receiver. To accomplish this, the attacker needs to be able to separate the signal $g(\mathbf{r}, t)$ into the individual terms of equation (1), that is

[2] for example via a real-time radio link that transmits the entire radio band used by the positioning system, shifted into another band

$$g(\mathbf{r},t) = \sum_i A_i \cdot g_i(\mathbf{r},t) + n(\mathbf{r},t) \tag{5}$$

with

$$g_i(\mathbf{r},t) = s_i\left(t - \frac{|\mathbf{x}_i - \mathbf{r}|}{c}\right) \quad . \tag{6}$$

This can then be reassembled into

$$g(\mathbf{r}', t - \mathit{\Delta} t) = \sum_i A_i \cdot g_i\left(\mathbf{r}, t + \frac{|\mathbf{x}_i - \mathbf{r}| - |\mathbf{x}_i - \mathbf{r}'|}{c} - \mathit{\Delta} t\right) + n'(t) \tag{7}$$

after choosing

$$\mathit{\Delta} t \geq \max_i\{|\mathbf{x}_i - \mathbf{r}| - |\mathbf{x}_i - \mathbf{r}'|\}/c \tag{8}$$

to preserve causality.[3]

4 Symmetric Security

The 24 orbiting satellites of the GPS constellation emit two separate broadcast signals $s_i(t)$, known as the C/A and Y signals. They both carry the same 50 bit/s data stream. It includes information on the current time and the exact orbital parameters of each satellite, which receivers need to calculate the time-dependent transmitter positions $\mathbf{x}_i(t)$. This data is transmitted using direct-sequence spread-spectrum (DSSS) modulation. The civilian C/A signal is modulated using a relatively short published spreading function. It can therefore not only be demodulated by the general public, but is also vulnerable to a signal-synthesis attack.

The military Y signal is produced by multiplying the 50 bit/s data signal with a secret and very long 10.23 MHz pseudo-random spreading sequence. This not only encrypts the signal like a stream cipher; it also spreads the 100 Hz mainlobe bandwidth of the data signal by a factor of 2×10^5 to 20 MHz. As a consequence, its peak power-spectral density is reduced by the same factor (53 dB) and ends up (according to [4]) roughly 28 dB below the thermal noise density seen by a typical receiver.

The original reason for this design were international regulations that protect microwave telephone links in the same frequency band from interference [4, p. 59]. Various tactical low-probability-of-intercept communication systems use DSSS modulation in a similar way to keep the power-spectral density of the transmission signal below the noise densities at expected eavesdropper sites.

[3] If the receiver forwards some unpredictable information received from each of the transmitters (for example their message-authentication codes) in real-time and the querying side has a means to verify these, then this creates another requirement for a selective-delay attack to succeed. At least four of the transmitters visible at the pretended position also have to be visible at the actual position. For GPS satellites (altitude: 20 200 km), this is usually the case within a few thousand kilometers.

In both the time and frequency domain, the Y signal disappears in the noise. Someone trying to manipulate the GPS Y code will therefore find it difficult to split $g(\mathbf{r}, t)$ up as in equation (5). As the shape of the waveforms is not known, correlation techniques cannot be applied to extract the phase of the Y signal from the noise.

It would therefore be very difficult to apply even a selective-delay attack on a GPS Y signal received with an omnidirectional antenna. The only option left to an attacker is to separate individual transmitters by using high-gain antennas. The use of at least four tracking dish antennas or a phased array may be feasible in some particularly well-funded attacks, but in most situations we would expect an attacker to be mobile and only be able to operate an omnidirectional antenna to capture $g(\mathbf{r}, t)$.

The problem with the GPS Y signal is of course that, since it is based on a single secret key, anyone in its possession can not only decode the Y signal to determine their position, but is also able to perform a signal-synthesis attack on any other Y-signal receiver. As a result, encrypted spread-spectrum navigation signals are so far used only in closed, mutually trusting user communities, in the case of the GPS Y signal the US military.

Another protection against signal-synthesis attacks has been proposed by Denning and MacDoran [2]. Their “location signature sensor” not only decodes the GPS C/A navigation signal in order to report its position to a remote authentication peer. It also detects and records a number of unpredictable attributes of the GPS signal, for example the clock noise added by the selective availability (SA) function of GPS to reduce the quality of service to the general public, as well as short-term fluctuations in the relative orbital positions that are not reported in the broadcast data. As long as the location signature sensors at both ends of the authenticated communication can see the same satellites, they can convince each other of being within a few thousand kilometers.

Again, this system only provides symmetric authentication and anyone able to verify the output of a location signature sensor in a geographical region will also be able to fake the output of such a sensor from anywhere within the same region.

5 Asymmetric Security

We now describe a new navigation-signal scheme that offers protection against signal-synthesis and selective-delay attacks comparable to that of an encrypted broadcast signal, that is one where the spreading sequence is a shared secret. However, the new scheme described in this section achieves this protection without the need to distribute and share any long-term secret keys among receivers. There is no information available to any receiver that would enable it to attack others. This approach is therefore particularly suited for open, international, civilian applications, where receivers are available in many forms to the general public and where some deployed receivers can be expected to be reverse engineered successfully by potential attackers.

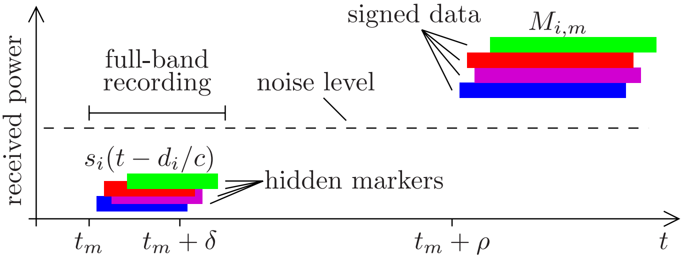

Fig. 2. In the proposed navigation-signal structure, first each transmitter X_i emits simultaneously from time t_m to $t_m + \delta$ its *hidden marker* $s_i(t)$. These pseudo-random waveforms overlap in the time and frequency domain. Their power is reduced significantly below the receiver noise level. The waveforms $s_i(t)$ are kept secret until time $t_m + \rho$ (typically a few seconds later). Then, signed information packets $M_{i,m}$ that describe the hidden markers are broadcast at normal power. Only after receiving these can receivers separate the markers from the recorded radio signal and determine their exact arrival times by detecting peaks in the cross-correlation function.

5.1 Hidden Markers

At regular preannounced times $t_1, t_2, \ldots$, for example every few seconds (or fractions thereof), all transmitters in the navigation system broadcast what we will call a *hidden marker*. We will discuss here only the transmission of hidden marker number m in this series, starting at system time t_m, understanding that this entire process will be repeated soon afterwards, starting at another time t_{m+1}, and so on.

The hidden marker is a rectangular pulse of duration δ, broadcast with DSSS modulation using a previously unpublished spreading sequence. Its power-spectral density is chosen such that it is at least 20 dB below the thermal noise when it arrives at the receiver. At the time at which this marker is transmitted, all the receivers and attackers can do is to digitize and buffer the entire antenna signal (filtered to the transmission band). This preserves in each receiver the information about the exact arrival time of the hidden marker, but it cannot be accessed yet. To determine this arrival time, the recorded noise has to be cross-correlated with the spreading sequence, in order to despread the marker and recover it from the noise.

However, the necessary information about the spreading sequence is not yet available at that time to any receiver. It is broadcast only after a delay ρ. Once this has been received, both regular receivers and attackers can identify and separate the markers in the recorded antenna signal. But any signal-synthesis or selective-delay attack can now be performed only with a delay $\Delta t > \rho$. By choosing ρ large enough, we can ensure that this delay can easily be detected by any receiver, even using an only loosely synchronized low-cost crystal clock. See also Fig. 2.

5.2 Transmitted Signal

In more detail, the steps taken at each broadcasting station X_i to generate the hidden-marker signal number m are:

1. Some time before t_m, X_i generates an unpredictable number $N_{i,m}$, for example using a cryptographically secure random-number generator.
2. This $N_{i,m}$ is used to seed a cryptographically secure pseudorandom bit-sequence generator $P(N_{i,m}, j) \in \{-1, +1\}$ that outputs a sequence of bits with indices $j = \{0, 1, 2, \ldots\}$.
3. From time t_m to $t_m + \delta$, X_i transmits the hidden marker, a sinusoidal carrier wave that is multiplied with the output of the seeded pseudorandom-bit generator, in order to spread its frequency spectrum[4]:

   $$s_i(t) = A \cdot \sin[2\pi f_\mathrm{c} \cdot (t - t_m)] \cdot P(N_{i,m}, \lfloor f_\mathrm{s} \cdot (t - t_m) \rfloor), \quad t_m \le t < t_m + \delta \tag{9}$$

   Here f_c is the chosen center frequency of the resulting signal and f_s is the bit rate of the spreading sequence, which is equivalent to half the mainlobe bandwidth of the resulting spectral power-density distribution

   $$|S(f)|^2 = (A/f_\mathrm{s})^2 \cdot \frac{\sin^2[\pi(f - f_\mathrm{c})/f_\mathrm{s}]}{[\pi(f - f_\mathrm{c})/f_\mathrm{s}]^2} \tag{10}$$

   The parameters t_m, f_c and f_s are identical for all transmitters (in other words, this is CDMA, not FDMA or TDMA), and the amplitude A is chosen low enough to bring the received signal well below the noise level.
4. At time $t_m + \rho$ (where $\rho \gg \delta$), X_i broadcasts a data packet of the form

   $$M_{i,m} = \mathrm{Sign}_{K^{-1}}\left[t_m, X_i, \mathbf{x}_i(t_m), N_{i,m}\right] \ , \tag{11}$$

   which is a message that is cryptographically signed with the private key K^{-1} of the navigation system and that reveals a full description of the previously transmitted hidden marker, including its transmission time t_m, the identifier X_i and exact location $\mathbf{x}_i(t_m)$ of the transmitter, and finally the unpredictable number $N_{i,m}$ used by that transmitter to spread the spectrum of this particular marker signal. Parts of this message may be transmitted earlier, as long as no information about $N_{i,m}$ is revealed until the nonce-release time $t_m + \rho$ has been reached.

5.3 Verification at the Receiver

By going through the following steps, each receiver R can use the hidden marker scheme to determine its position in a way that is robust against signal-synthesis and selective-delay attacks:

[4] We use here binary phase-shift keying (BPSK) to modulate the hidden marker signal, but many other modulation schemes could be used equally, including the binary offset carrier (BOC) modulation techniques used in some more recent navigation systems.

1. The implementation of the receiver's local clock $t_R(t)$ must not be influenced in any way by information received through navigation signals. We assume that it has a known maximum relative frequency error ε_f, such that

   $$\left|\frac{t_R(t+\tau)-t_R(t)}{\tau}\right| \le \varepsilon_\mathrm{f} \ .$$

   We also assume that t_R was last adjusted by an authenticated two-way clock synchronization from a trusted source at system time $\hat{t}$ such that $|t_R(\hat{t})-\hat{t}| \le \varepsilon_\mathrm{s}$. The error $u_R(t)$ of the local clock $t_R(t)$ is then bounded by

   $$|u_R(t)| \le \varepsilon_\mathrm{f} \cdot (t-\hat{t}) + \varepsilon_\mathrm{s}, \quad \text{for } t \ge \hat{t} \ . \tag{12}$$

   Simple crystal oscillators offer $\varepsilon_\mathrm{f} < 10^{-5}$ and authenticated two-way clock synchronization over wireless computer networks usually offers $\varepsilon_\mathrm{s} < 100$ ms.
2. During a time interval slightly larger than $[t_m, t_m+\delta]$, the receiver digitizes the entire frequency band $[f_\mathrm{c}-f_\mathrm{s}, f_\mathrm{c}+f_\mathrm{s}]$ with a sampling rate of at least $4f_\mathrm{s}$ and stores it in a RAM buffer $B(t_R)$.
3. It then waits for the arrival of the broadcast messages $M_{i,m}$ and discards those whose signature cannot be verified using the navigation system's well-known public key K or whose marker time t_m does not match the marker time for which the receiver initiated the wide-band recording in the previous step.
4. For each $N_{i,m}$ extracted from a message $M_{i,m}$ that passed these checks, the receiver now generates the spreading sequence $s_i(t_R)$ from equation (9). These are then cross-correlated with the RAM buffer B[5]:

   $$C_{i,m}(\tau) = \int_t B(t) \cdot s_i(t+\tau)\,\mathrm{d}t \tag{13}$$

5. For each cross-correlation result $C_{i,m}$, the position $\hat{\tau}_{i,m}$ of the largest peak in it is recorded, together with the relative amplitude $w_{i,m}$ of any second-largest peak.
6. Of the recorded tuples $(i, \hat{\tau}_{i,m}, w_{i,m})$ the receiver now discards all where the second-largest peak is not attenuated by at least a configurable security factor W relative to the largest peak. (The reason for this step will become clear in Sect. 6.)
7. The remaining peak-positions $\hat{\tau}_{i,m}$ are then used as pseudoranges

   $$\tilde{d}_i = c \cdot \hat{\tau}_{i,m} = |\mathbf{x}_i - \mathbf{r}| - c \cdot u_R \tag{14}$$

   and the resulting set of equations, which use the received digitally signed transmitter positions $\mathbf{x}_i$, is solved for $\mathbf{r}$ and u_R.
8. The result is accepted, if the u_R value remains within the clock uncertainty allowed by inequality (12) and is smaller than the time delay ρ for the publication of the spreading-sequence seed values.

[5] In a practical implementation, recording and cross-correlating the hidden marker may be done after conversion from f_c down to a lower intermediate frequency.

This scheme utilizes the fact that there are now low-cost analog-to-digital converters available, with sampling rates of more than 100 MHz. This, together with falling RAM prices, has made it feasible to record in battery-operated low-cost devices at an intermediate frequency of up to 50 MHz for several seconds entire RF bands that are 20 MHz or more wide, as they are occupied by the GPS Y signal.

5.4 Optimized Broadcast Data

Existing navigation systems operate with comparatively low bit rates for the transmission of data (e.g., 50 bit/s for GPS). Therefore, a concern may be the length of the cryptographically authenticated message $M_{i,m}$, which releases the number $N_{i,m}$ and binds it securely to the transmission parameters of the corresponding hidden marker. A digital signature alone consists of several hundred bits, so the length of $M_{i,m}$ might become a limiting factor for the rate at which hidden markers can be transmitted. Fortunately, there are several optimizations of the scheme possible, which reduce the required bit rate.

The individual messages $M_{i,m}$ can be consolidated into a single system-wide message M_m. In particular, M_m could contain only a single unpredictable number N_m, from which then the individual seed numbers $N_{i,m} = g(N_m, i)$ can be derived in a predictable way. The function g could be something as simple as addition. Individual transmitters can also vary the order in which they transmit the elements of M_m, such that receivers can compile the complete M_m faster from the parallel reception of several transmitters than from listening to merely a single one.

Instead of including N_m in M_m as a separate data field, it could also be derived from M_m's digital signature, which is already unpredictable. The transmitters would then have to commit to the content of M_m before time t_m is reached, and would lose their ability to update position and time using the latest measurements, in return for eliminating the need to transmit N_m. Where the values of t_m and $\mathbf{x}(t_m)$ can be predicted well in advance, only the marker serial number m itself needs to be signed in each M_m. The parameters for predicting t_m and $\mathbf{x}(t_m)$ from m can then be broadcast as a separate message much less frequently.

Alternatively, it is also possible to avoid the addition of a digital signature to each M_m entirely by using a symmetric *stream-authentication method*, such as the one proposed in [5]. Such schemes operate on a principle very similar to the hidden-marker system presented here. They replace the digital signature with a symmetric message-authentication code, and release the – for the receiver unpredictable – authentication key only after a delay (equivalent to ρ above) that is longer than the clock uncertainty of the receiver. Only the first message in such a stream needs to be digitally signed. The message-authentication keys used in all further packets are derived from their respective successor, using a secure one-way function. They can therefore be verified from their respective predecessor, as soon as they are released.

If we used a standard stream-authentication method, such as [5], directly to protect the messages M_m, the authenticity of the hidden marker could only be verified after *two* delay periods ρ, one to protect N_m and the other to protect the message-authentication key. This problem can be avoided by eliminating the message-authentication code, and instead making all the values N_m directly parts of a one-way chain.

In more detail, here is how we can combine the hidden markers number $m_0, m_0 + 1, \dots, m_0 + n$ into a single marker stream that requires only one single digital signature:

1. The transmitters pick at random an unpredictable final number N_{m_0+n} for the stream, and then generate a number N_m for each of the n previous markers, via the recursion $N_m = h(N_{m+1})$ (for $m_0 \leq m < m_0 + n$). This way, the first number will be $N_{m_0} = h^n(N_{m_0+n})$. Here, h is a secure one-way function, that is a function for which, given a value y, it is computationally infeasible to find a preimage x with $h(x) = y$.
2. The transmitters then broadcast some time before t_{m_0} the message
$$M_{m_0} = \mathrm{Sign}_{K^{-1}}\left[m_0, h(N_{m_0}), D\right] \quad , \tag{15}$$
where D is a parameter set that describes how the values t_m and $\mathbf{x}_i(t_m)$ can be calculated from a given station number i and marker number m.
3. Finally, the transmitters broadcast from time t_m to $t_m + \delta$ their respective hidden markers, generated from $N_{i,m} = g(N_m, i)$, and they broadcast at time $t_m + \rho$ the message N_m, and this for each $m \in \{m_0, \dots, m_0 + n\}$, as described in Sect. 5.2.

The receivers follow the same steps as described in Sect. 5.3, except that a digital signature is now verified only for the first message M_{m_0} in each stream. The subsequently released value N_{m_0} is verified against the signed value $h(N_{m_0})$ in M_{m_0}. All the subsequently released values N_m (for $m_0 < m \leq m_0 + n$) are then verified with the test $h(N_m) = N_{m-1}$. The parameters t_m and $\mathbf{x}_i(t_m)$ are calculated from the signed parameter set D (which in a satellite navigation system, for example, would include the orbital parameters).

This way, apart from the signed message M_{m_0} that precedes a stream of $n+1$ consecutive markers, only a single number N_m needs to be broadcast per marker. It will not have to be longer than 60–80 bits in practice, just enough bits to make a brute-force inversion of h infeasible within the time interval $t_{m+1} - t_m$.

The length $n + 1$ of these marker streams is limited by the requirement that newly activated receivers, and those that missed one of the values N_m, should not have to wait long until they can restart the authentication chain with the start of a new stream.

6 Selective-Delay Using High-Gain Antennas

There is an alternative way of separating the right side of equation (5) into the terms contributed by the individual transmitters, which does not depend on knowing the spreading functions. If the approximate positions of transmitters are predictable, at least four of them can be targeted with directional antennas.

If the gain of these antennas is high enough to lift the broadcast signals out of the background noise, demodulation and threshold operations can be applied in order to free the signal of one station completely from any interference by the others, enabling a selective-delay attack that cannot be detected. The only protection against this attack appears to be to keep the signal strength enough below the noise limit to require antennas so large that their use during a practical attack becomes infeasible.

If the signal-to-noise ratio achievable with directional antennas is not sufficient for separating and decoding the signals directly, then the attacker can still delay the raw antenna signals and mix them together for the receiver. In practice, no directional antenna will be able to suppress the signals from all other transmitters completely. This will cause weaker shadow peaks to show up in the cross-correlation results for each transmitter station, picked up and contributed by an antenna pointing to another station, at the relative delay applied there. The security parameter W in the receiver algorithm from the previous section defines, how sensitively the receiver should react to such shadow peaks. This sensitivity could be made dependent on the distance in time from the main peak, such that a selective-delay attack with directional antennas is not confused with secondary peaks caused by plausible multi-path propagation.

7 Example Parameters

The technique presented in Sect. 5 is particularly suited for navigation systems that transmit from medium-earth-orbit (MEO) satellites, such as GPS, Glonass or Galileo. In this setting, there are clear lower and upper bounds for the ranges between receivers and visible transmitters (e.g., 20 000–26 000 km for GPS), which helps to ensure a uniform received signal strength, at least outdoors. The transmitters also move fast enough to complicate the use of directional antennas. For other types of pseudo-range navigation systems, such as land-based longwave transmitters (e.g., LORAN-C) or short-range ultrasonic or ultra-wideband-radio positioning systems, more complex schemes may be needed that involve hidden markers broadcast at a wide range of power levels.

The security of the scheme is based on the assumption that at any receiver position, the time intervals during which hidden markers arrive from the various transmitters will overlap substantially. With MEO transmitters, ranges can vary by up to 6000 km. This corresponds to 20 light milliseconds, and the duration of the hidden marker will have to be at least one or two orders of magnitude longer than that. A typical value may be $\delta = 1$ s.

We need to chose the signal strength, such that a clear peak appears after the cross-correlation with the correct spreading sequence in a receiver, while keeping on the other hand the power spectral density of the broadcast signal well below thermal noise. Integrating during a cross-correlation for an entire second is roughly equivalent to filtering the noise bandwidth of a signal down to 1 Hz.

As a very simple example, if we quantify (pessimistically) the thermal background noise to be expected by a receiver with an equivalent antenna temperature of 290 K (including atmospheric noise, cosmic background radiation, antenna temperature noise, transmission line losses, amplifier noise [4,6]), this corresponds (after multiplication with the cross-correlation bandwidth of 1 Hz and Boltzmann's constant) to a noise power level of about −204 dBW. If the transmission power of each hidden marker is selected such that about −170 dBW reach the receiver, then the 34 dB signal-to-noise ratio obtained this way ensures that spurious peaks in the cross-correlation output caused by noise will remain much smaller than the peak caused by the hidden marker.

If we use, as the GPS Y-code does, a spreading frequency of $f_\mathrm{s} = 10$ MHz, then an attacker who does not know yet the spreading sequence will have to work with the full 20 MHz mainlobe bandwidth of the broadcast signal. Even with a much better omnidirectional antenna, with an equivalent noise temperature of only 100 K, this still leaves −136 dBW received noise power, which is 34 dB above the signal energy and therefore will render the broadcast signal unrecognizable.

A 20 MHz wide intermediate frequency signal can be recorded comfortably with a sampling frequency of 200 MHz. With a signal-to-noise ratio of −34 dB, there is little point in storing more than one or two bits per sample after analog-to-digital conversion, as the quantization noise would still be small compared to the thermal noise. Therefore, the entire hidden marker can be practically stored in not more than 25 MB of RAM.

The choice for the delay time ρ after which the information about the spreading sequence is released depends on how frequently a receiver is assumed to get in contact with a trusted source of the system time t, and how stable its local clock is. If we take as an example a maximum time between resynchronizations of $t - \hat{t} < 1$ week, a local clock frequency error of $\varepsilon_\mathrm{f} < 10^{-5}$, and a synchronization error of $\varepsilon_\mathrm{s} < 1$ s, then from equation (12), $\rho = 10$ s $> u_R$ would appear to be a suitable choice. Where no single value for ρ can be found that suits all applications, it is possible to broadcast hidden markers with a range of different time delays.

8 Conclusions

This paper considered an aspect of the security of pseudoranging positioning systems, such as GPS, namely how a receiver can be misled about the position of its antenna if an attacker is allowed to insert a signal-manipulation device between the receiver and the antenna. We have shown that positioning systems currently offer no defense against signal-synthesis or selective-delay attacks without the receiver obtaining all the information necessary to mount these attacks on others.

We outlined a new signal structure and the corresponding verification algorithm for receivers that solves this problem. A weak spread-spectrum broadcast signal is temporarily hidden in background noise while receivers buffer the entire radio band in RAM. The despreading key is only published after a time that is larger than the uncertainty of the local clock in the receiver, at which time both a signal-synthesis and a selective-delay attack can easily be detected. Such keys can be authenticated efficiently by making them part of a one-way chain.

The system is still based on the pseudoranging principle and uses only a low-cost local clock in the receiver. It can therefore still be defeated by relaying attacks. Against these, we see no solution other than using a more expensive highly-stable oscillator in the receiver, or using authenticated two-way ranging, both of which would be able to detect the added delay.

The system is also vulnerable to selective-delay attacks involving at least four high-gain directional antennas. A security parameter that limits the height of shadow peaks in the cross-correlation result can be used to control the minimum antenna gain needed for this attack to succeed, thereby limiting its practicality.

References

1. Paul Kallender: Omron uses GPS to catch a car thief. EE Times, 12 June 2001. http://www.eetimes.com/at/news/OEG20010612S0059
2. Dorothy E. Denning, Peter F. MacDoran: Location-based authentication: Grounding cyberspace for better security. Computer Fraud & Security, Elsevier, February 1996, pp. 12–16.
   http://www.cosc.georgetown.edu/~denning/infosec/Grounding.txt
3. Electronic tagging: A virtual prison? BBC News Online, 7 January, 2000.
   http://news.bbc.co.uk/1/hi/special_report/1999/02/99/e-cyclopedia/594314.stm
4. J.J. Spilker Jr.: GPS signal structure and theoretical performance. In B.W. Parkinson and J.J. Spilker Jr.: Global Positioning System: Theory and Applications – Volume I, Progress in Astronautics and Aeronautics, Volume 163, American Institute of Aeronautics and Astronautics, Washington DC, 1996, ISBN 1-56347-106-X.
5. Adrian Perrig, Ran Canetti, J.D. Tygar, Dawn Song: The TESLA broadcast authentication protocol. CryptoBytes, Vol. 5, No. 2, pp. 2–13, RSA Laboratories, Summer/Fall 2002.
6. Radio noise. Recommendation ITU-R P.372-7, International Telecommunication Union, Geneva, 2001.