Signal Authentication in Trusted Satellite Navigation Receivers

Markus G. Kuhn

Abstract In some security-critical applications, a GPS satellite-navigation receiver is integrated with a tamper-resistant cryptographic module in order to provide remote attestation of location. Those in possession of the receiver may have an interest in it producing an incorrect output. Vehicle and container tracking, usage-based road charging, prisoner tagging, location-based access control are just some secure-positioning examples where anti-spoof measures against local attackers are of concern. The discussed tests and trust metrics can help a tamper-resistant navigation-signal receiver to distinguish authentic signals at the RF antenna port from those forged using a signal simulator.

Key words: global navigation satellite system, GNSS, global positioning system, tamper-resistant GPS receiver, signal authenticity, anti-spoof measures, location-based service

1 Introduction

Physical location can be an important security parameter, whether for location-based access control or to audit the whereabouts of goods and people. In outdoor applications, location is often most easily determined with a global navigation satellite system (GNSS) receiver. This means today primarily GPS [1, 2], but the list is growing (GLONASS, Galileo, Beidou/Compass, ...). Each of these operates a constellation of Earth-orbiting satellites that broadcast a high-precision time signal, along with a low-bitrate data stream (50–1000 bit/s) that carries orbital-position (ephemeris) predictions and calibration data. Receivers measure the time-of-arrival differences of at least four satellite signals and then solve a system of equations to determine both

Markus G. Kuhn
Computer Laboratory, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB3 0FD, e-mail: Markus.Kuhn@cl.cam.ac.uk

their current location and time, with accuracies of a few meters and tens of nanoseconds. Even higher accuracies can be achieved by using nearby reference receivers for calibration.

GNSS receivers may be integrated with tamper-resistant cryptographic modules, for security-critical applications where the person in possession of the device has an interest in it misreporting its location. The purpose of such devices may be to attest their current location to a remote observer, via an authentication protocol. It may also attest the device's recent location and velocity history via an authenticated recording, or enable some functionality based on its location. Potential application examples include

- anti-theft tracking systems for vehicles and transport containers, which automatically alert the owner if a vehicle no longer follows its expected route;
- prisoner-tagging systems that permit probation officers to remotely monitor curfews and probation conditions;
- road tax and insurance fees for motor vehicles calculated in on-board units based on actual usage, with algorithms that incorporate information about speed, route and travel times in order to take into account externalities and risk;
- road speed limits enforced electronically using on-board navigation systems that determine the current location, look up the local speed limit, and communicate that to both driver and engine controller.

Some of these applications are already deployed, others may well evolve from existing road-usage or congestion charging systems, or the tachographs or speed limiters found already in many commercial vehicles.

The design of tamper-resistant embedded computers is already on its way to become a well-understood engineering discipline [3, 4], supported by a range of commercially available components, such as intrusion sensors [5], battery-backed RAM key storage with emergency zeroization mechanisms [6], shielding against compromising emanations and other side-channel countermeasures.

Therefore, the focus is here on the other main vulnerability of tamper-resistant GNSS receivers: that their antenna input could be fed with a simulated signal rather than from the satellites. A specialized portable signal generator could synthesize a GNSS antenna signal that causes the receiver to report an alternative position, velocity or time to the connected cryptographic module. A simple example would be a device that records the route taken by a lorry and then replays the coordinates slower to a GNSS signal simulator that the driver has installed to replace the satellite reception antenna, such that a speed-limiter function is not triggered, but the tachograph still shows a realistic-looking record of the driven route. (Police have already uncovered similar manipulations of speed-sensor signals in existing tachographs [7].)

What measures could a GNSS receiver implement to assess the authenticity of a received GNSS signal and the resulting navigation solution? It is important to note that the notion of authenticity of a GNSS signal goes beyond the usual meaning of message authenticity in cryptographic protocols. We have to protect not only the authenticity of the transmitted navigation data, but also that of the relative arrival times (the *pseudo ranges*) of the transmitted spread-spectrum waveforms, within

better than a microsecond. Both together form the basis for calculating the navigation solution.

1.1 Environmental assumptions

This discussion focuses in particular on tamper-resistant GNSS receivers (also called trusted receivers) that are assumed to be in the hands of the adversary. We make the following assumptions:

The trusted receiver consists of an antenna, a circuit for demodulating and tracking the GNSS signal, and a secure microcontroller. The microcontroller stores cryptographic secret keys and uses these to attest (e.g., by time stamps, digital signatures, or similar cryptographic protocols) to a remote party the current location (or the recent location history).

The receiver's RF front-end, signal processing circuitry, local oscillator, and the secure microcontroller are all enclosed in a tamper-responding shield that the adversary is unable to penetrate without destroying the secret keys stored inside.

This tamper-responding enclosure is equally securely attached to the object whose location is ultimately of interest (car, laptop, etc.). (This attachment could be secured, for example, by strong mechanical bonding, by detachment sensors, or by some cryptographic distance-bounding protocol to another tamper-resistant CPU in the monitored object.)

The adversary has full control over the RF signal received by the device, and in particular, may disconnect the antenna and connect the receiver's RF input instead to a signal generator programmed to emulate GNSS broadcast signals, with the aim to cause the secure microcontroller to process fake position information. Alternatively, where the antenna is not easily detachable, the adversary may also place the tamper-resistant receiver's antenna inside a shielded enclosure, along with transmission antennas connected to a signal generator.

The first commercially available GNSS simulators have been very expensive and specialized devices. However, during the past decade, numerous low-cost components (high-speed DACs, FPGAs, DSPs) and standardized platforms for building software-defined radio applications (GNU Radio project, Ettus USRP, various DSP processor/FPGA evaluation boards, etc.) have become available. This makes it practical now to design high-quality GNSS signal simulator prototypes with a hardware budget in the region of 1–2 k$ [8]. The result of such design efforts can easily be shared as open-source software, which will substantially increase the number of people able to understand, implement and customize such devices. With the increased availability of GNSS simulation capabilities, attacks involving GNSS signal simulators should be expected as soon as attractive targets emerge, namely mass market applications that involve remote-attestation GNSS receivers (e.g. location-based access control, pay-as-you-drive road charging systems) where the holder of the trusted receiver has an incentive to spoof its input signal.

1.2 Related technologies

Much of the existing literature on GNSS signal spoofing and jamming has focused on a remote attacker scenario, where the receiver is believed to be in the hands of a user who is interested in it finding a correct navigation solution (e.g., a soldier) and where the antenna is still exposed to genuine GNSS broadcast signals. A remote attacker can only add additional signals to the receiver's RF environment from a distance. Anti-jamming and anti-spoofing countermeasures aim to suppress these, to preserve the availability of the (also present) genuine signals. Examples for such countermeasures include

- the use of directional antennas (beam-forming networks) to suppress unusually strong GNSS signals, which are unlikely to be coming from a genuine satellite;
- an adaptive filter for suppressing interfering narrowband signals;
- the combination of two tracking circuits, where the job of the first is to track the spoofed (often stronger) signal, such that it can be subtracted from the input in order to allow the second tracking circuit to follow the remaining, weaker genuine signal.

In contrast, we assume here a local attacker who can easily suppress any trace of the genuine GNSS signal at the RF input during an attack, where the entire antenna signal may be fake. Rather than looking for traces of a weak genuine signal in the presence of a stronger spoof signal, the signal authenticity mechanisms discussed in Section 2 below focus on discovering signal characteristics that help to distinguish between a genuine and a simulated GNSS antenna input.

1.3 Goals

Any practical GNSS signal simulator will produce an idealized signal that lacks some of the subtle characteristics found in a genuine signal. Ultimately, any mechanism for assessing signal authenticity can only be effective if the receiver's designer uses a more accurate model of a genuine signal than the signal simulator's designer. With enough effort and resources, any of the methods discussed in Section 2 can be circumvented, either by carefully emulating all the tested characteristics, or by appropriately modifying a genuine signal. However, such a simulator may not in practice be an attractive means of defeating a given security application (operating cost, mobility, physical dimensions, etc.). Also, if it were openly sold, it would have to provide capabilities substantially beyond the type of simulators normally used legitimately for the development, testing and maintenance of GNSS receivers. Therefore, it could be identified as having been specifically designed to circumvent the proposed security mechanisms for assessing signal authenticity, and its sale might be illegal under existing cybercrime legislation.

We can distinguish between two broad categories of methods for assessing signal authenticity:

- **Instant methods** assess signal authenticity almost as soon as a navigation solution has been found, and should not extend the duration of the normal lock-on process by more than a few seconds. They are of particular interest where an action (such as a network login with location-based access control) has to be blocked instantly if there are substantial doubts regarding the authenticity of the navigation solution.
- **Cumulative methods** monitor the GNSS signal over many hours or days and report in the end whether there has been substantial evidence of a fake signal during this period. Such methods may be applicable in accounting applications (e.g., road charging), where the damage that a successful attacker can cause is proportional to the time that the fake signal is accepted by the receiver. Where cumulative methods can be used, a wider range of verification techniques is available.

  Applications that can rely on cumulative methods have another security advantage: the detection of a simulated signal need not be made known to the user of the receiver immediately, and therefore deprives an adversary from the rapid feedback that helps optimizing a signal generator. Instead, the signal-authenticity assessment can just be silently recorded, helping the operator of the protected application to estimate the level and nature of attacks taking place and to focus investigation and countermeasures appropriately.

2 Techniques

2.1 Secret spreading sequences

GPS and planned GNSSs use direct-sequence spread-spectrum modulation in their broadcast signals. The low-bitrate (< 1 kbit/s) data signal is XORed with a high-bitrate (> 1 Mbit/s) pseudo-random spreading sequence, before the result is used to modulate the phase of a carrier sine wave.

A range of possible techniques rely on the fact that GPS broadcasts both its civilian (C/A) and military (Y) signals at a power-spectral density substantially below the background noise level. Receivers with omnidirectional antennas are therefore unable to decode the individual "chip" symbols of the spreading sequences and can only detect a cross-correlation with a known sequence of at least a few hundred chips. In addition, the GPS Y signal is, due to its encryption, not predictable by non-military users and therefore difficult to reproduce in a simulator. Galileo is foreseen to broadcast similar weak broadband signals and to provide a similarly encrypted (and therefore for most users unpredictable) signal in its Public Regulated Service (PRS), although the details have yet to be finalized and published.

2.1.1 Conditional access

One possibility for assuring signal authenticity is, of course, to keep the spreading sequence (ranging codes) used secret and non-repeating. A conditional-access system, similar to those already widely implemented in the direct-broadcast satellite pay-TV industry, has to ensure that the cryptographic keys needed to predict the spreading sequence for the near future are distributed to tamper-resistant modules handed out to authorized subscribers. Such modules then determine, based on received entitlement management messages, to which level of service the user of each module is entitled, and then extract from also received entitlement control messages the necessary cryptographic keys for accessing these services. The U.S. military uses already a form of conditional access for the encrypted Y signal, and subscriber modules appear to be planned for the Galileo commercial, safety-of-life and military services.

The tamper-resistant subscriber modules have to be very carefully designed such that they cannot be abused by a spoofing attacker as a component of a signal generator that can predict the secret sequences about to be broadcast. This would typically involve performing the correlation and tracking operation inside the module, such that the keys used to generate the next parts of the spreading sequence never leave the tamper-resistant envelope. Another challenge, which has already been studied in detail over the past 20 years in the context of pay-TV conditional access systems, is to design a broadcast-encryption and traitor-tracing key-management system that can recover its security after a small number of subscriber modules have been broken [9]. This is not an easy task if the available broadcast data channel has only a low bit rate.

If these aspects can be secured, the main option remaining to an attacker interested in simulating a conditional-access signal is to use tracking high-gain antennas. These could improve the signal-to-noise ratio (SNR) to a level that allows reliably detecting the individual chip symbols in the broadcast spreading sequences in real time. The attacker can then slightly delay and remix them in the signal generator (selective delay attack with high-gain antennas, see [10]) to simulate how they appear relative to each other at the pretended location. Such attacks can be made more cumbersome in two ways:

- Keep the broadcast power density well below the background noise level, in order to maximize the physical antenna dimensions required (e.g., large parabolic dish or long helical antennas).
- Keep the symbol rate high, in order to make it more difficult for an attacker to forward the signals received at a stationary set of directional antennas to a mobile signal generator.

2.1.2 Delayed release of spreading sequences

A method to achieve similar signal-integrity assurance as a conditional-access system can provide, but without the overhead and risk of compromise of a tamper-

resistant subscriber module and associated key-distribution infrastructure, was proposed independently by Scott [11] and Kuhn [10]. The idea is that the spreading sequences used are secret at the time of their broadcast, but information to reconstruct them is broadcast with a delay of a few seconds. This allows tamper-resistant receivers to discover, with a short delay, the genuine broadcast signals, using FFT-based cross correlation on recorded segments of the entire transmission band. At the same time, this forces the designer of a signal generator to delay the signal also by a few seconds, which an independently synchronized UTC clock in the receiver can easily detect.

The delayed release of the spreading sequence remains the most practical and resilient single integrity assurance method currently known:

- It does not require support from a network of reference stations.
- It does not rely on the security of a subscriber key-distribution infrastructure.

Such a scheme could be piggybacked on top of an existing service that broadcasts using a secret spreading sequence. The latter would have to be generated by a pseudo-random-bit-sequence (PRBS) generator that is seeded with a new secret start value in regular intervals. The satellites then simply would have to occasionally broadcast a subset of the PRBS seeds that have been used, but with some delay. For example, the encryption scheme for the Galileo PRS signal could be designed such that keys that generate only short intervals of the spreading sequence can be released without affecting the security of its conditional-access users. The key K obtained by the conditional-access modules would in such a system not be applied directly to generate the spreading sequence. Instead, it would be used to encrypt a timestamp t that identifies a short time interval (e.g., 1 second) in order to obtain a short-term intermediate key $K_t = E_K(t)$ which is then used to seed the PRBS generator that generates the actual pseudo-random-noise (PRN) spreading sequence, one second at a time. E is some suitable keyed pseudo-random function, e.g. a cipher, message-authentication or secure-hash function. A small subset of the short-term intermediate keys K_t is then released with a short delay. The interval length (e.g., one second), the subset of the released intermediate short-term keys K_t (e.g., one K_t every 20 seconds) and the delay (e.g., 10 seconds) has to be chosen such that their publication does not enable practical spoofing of regular receivers of the conditional-accesss service, whose tracking loops would have to be designed to be immune to regular but brief bursts of old spreading sequence.

2.1.3 Permanently secret sequences

Can we adapt the basic idea from the preceding Section 2.1.2 if we have users who never get access to the spreading sequence used by the satellites? This is the case, for example, for civilian users regarding the GPS Y code. A reference station still can record the spreading sequence, but has to use high-gain antennas that lift the SNR sufficiently to allow it to receive and detect the spreading sequence directly, convert it into a bit stream (10.23 Mbit/s for the GPS Y code) and arrange for that to

be forwarded to the tamper-resistant receiver, who will correlate it with a pre-agreed brief concurrent recording of the full transmission band. This method, discussed in more detail by Psiaki [12], works similarly as the one outlined in the preceding Section 2.1.2, but is more expensive to implement:

- It requires a reference station with large tracking antennas (ideally at least four, fewer if only probabilistic verification is required).
- It requires a higher-bandwidth secure communication link to the tamper-resistant receiver. Entire spreading sequences received during the pre-agreed time window will have to be provided to the receiver for delayed cross-correlation there, rather than just K_t seed values that generate them. (Transmitting the received and signed raw spectrum during remote attestation from the tamper-resistant receiver is another option, but requires an even higher bit-rate communications channel.)

2.2 Individual receiver antenna characteristics

2.2.1 Directional characteristics

If the receiver antenna is installed at a fixed location, or mounted on a car, the receiver might be able to observe the directional variation of amplitude (and perhaps even phase?) as the satellites move along the sky with known azimuth and elevation. Mounted on a car, the orientation of the antenna will normally only vary in azimuth (yaw), due to curves, and to a limited degree in elevation (pitch), due to hills. Both angles can there be inferred from the velocity vector determined by the receiver, assuming that roll movements are very limited and temporary. In particular, the yaw motions of a car will cause the satellites to quickly scan a substantial part of the directional characteristic of the antenna, which can be monitored for changes.

A receiver could characterize the directional characteristic of its antenna from the received signal strength, in particular if data about absolute signal strength from the automatic gain controller and the correlator is available. The designer of the trusted receiver could chose an antenna type specifically for its structurally rich directional pattern, for instance a fractal antenna rather than a simple dipole, and could even individually vary the exact antenna shape and encase it in opaque resin, in order to increase the effort needed by an attacker to recognize and model its characteristic. While an attacker could measure the individual antenna pattern of each replaced antenna, and program the signal simulator accordingly, this adds substantially to the effort needed to implement an attack, ideally beyond being economically attractive for a mass-market fraud device.

2.2.2 Impedance test

Where a custom RF frontend is being designed for a tamper-resistant receiver, this opens the possibility to add circuitry that characterizes the frequency-dependent

impedance of the connected antenna occasionally, raising an alarm if that changes substantially. Possible techniques include time-domain reflectometry, VSWR measurement (if there is a transmission line), or vector network analysis. Especially if the antenna has been produced deliberately with characteristic invisible manufacturing variations, the need to keep the antenna attached (i.e., use a shielded enclosure around it), or to emulate the antenna impedance, represents a substantial complication in the appropriate connection of a signal generator, possibly one that makes mass-market sale of signal generators far less feasible.

2.3 Consistency with reference receivers

One group of signal-authenticity measures compares characteristics of the received GNSS signals with the same characteristics measured at the same time by a network of trusted reference receivers. These reference receivers can either be dedicated stations, secured by traditional anti-jamming measures (e.g., distance, directional antennas), or they can be obtained by assuming that the majority of the signal characteristics reported by a fleet of trusted receivers is genuine, allowing outlier detection.

2.3.1 Time

All existing or planned GNSSs broadcast Coordinated Universal Time (UTC) with an accuracy better than a microsecond. More accurate signal-simulation techniques often involve incorporating data from reference stations, and this usually requires delaying the generated signal. Therefore, a trusted GNSS receiver should first of all verify the UTC received by GNSS with an independent authenticated source of UTC. This can be accomplished by operating a local UTC clock, independent from any received GNSS signal. This clock should be synchronized regularly via an authenticated challenge-response time protocol, like NTP. Such network time protocols can, depending on the communication link, achieve UTC accuracies of a few tens of milliseconds or better. The resulting clock accuracy is mostly a function of how frequent these phase and frequency adjustments can be made compared to the undisciplined frequency stability of the local oscillator. If the received UTC(GNSS) differs from the UTC(NTP) in the local clock by substantially more than the latter's uncertainty (e.g., a few tens of milliseconds), a clear indication has been found that either the GNSS signal or the independent source of UTC has been manipulated.

2.3.2 Navigation data

The independently synchronized UTC clock in a trusted receiver can also be used to timestamp a revision history of the navigation messages received from individ-

ual satellites. This revision history, which records (with a resolution of a few tens of milliseconds) when which bit in the navigation message has been observed to change, can then be compared with the corresponding revision history collected by the reference receivers.

Bypassing this measure would require an attacker to either be able to anticipate the content of navigation messages that are newly uploaded into satellites, or implement in the signal generator a specialized receiver that provides a real-time feed of the navigation signal. The proprietary binary protocols of typical existing GPS receiver chipsets output changes to navigation messages only with significant delay, usually awaiting the completion of frames and parity checks. If a signal simulator is merely fed with such delayed navigation-message updates, its use would be detected by this measure.

This test relies on the satellite operator not publishing all updates to the broadcast data in advance. It is the more effective the more frequent the navigation data changes in unpredictable ways. For this reason, designers of future GNSS signals could add to navigation messages unpredictable random bits, such as time-dependent message authentication codes or hash chains.

2.3.3 Pseudo-ranges

If more than four satellites are in view simultaneously, an over-determined system of equations will lead to the navigation solution. Satellite clock and ephemeris errors, as well as atmospheric path delays, will then cause inconsistencies, usually of several meters. A tamper-resistant receiver with access to raw pseudo-range measurements could compare these inconsistencies with those observed by a nearby reference receiver. Inconsistencies caused by the atmosphere will vary geographically, and therefore would force the adversary to have access to a reference receiver in the vicinity of the emulated location. (Experience with differential GPS suggests that pseudo-range inconsistencies show a substantial loss of correlation at distances larger than a few tens of kilometres.)

There are regional networks of differential GPS stations that publish pseudo-range inconsistencies[1] that both the trusted receiver and the adversary could refer to. However, as long as they publish their information only with a delay larger than the auto-correlation width of the data, they could be used for verifying pseudo-range inaccuracies without enabling a signal generator to simulate them in real time.

[1] e.g., the OS Net RINEX data server available on `http://gps.ordnancesurvey.co.uk/` for the British Isles, or (continent-wide, at currently much lower station density) the data from augmentation services such as EGNOS/SISNeT

2.4 Receiver-internal plausibility tests

Beyond the minimally necessary processing needed to achieve a navigation solution, receivers can implement additional consistency checks without requiring a connection to a network of reference receivers. A number of such tests have been proposed and are often referred to in the literature as receiver autonomous integrity measures (RAIM). They were originally aimed primarily at detecting accidental malfunctions in the GNSS, such as one of the satellites suffering from a phase jump or frequency deviation in its local oscillator, or the broadcast of incorrect or out-of-date navigation messages. They have also been proposed to detect very simple types of GNSS signal simulation [13] and would force the attacker to use a more complete simulation model, including realistic and up to date navigation data and parameters.

2.4.1 Elevation limit

A very simple check involves verifying that each satellite from which a signal is received actually claims to be above the horizon at the moment. This test was proposed by [13] to detect if a very simple type of signal simulator is used that always transmits a fixed number of satellite signals (e.g., 10), even if their simulated position is well below the horizon. This test can be implemented with many consumer receivers, which output the azimuth and elevation of all tracked satellites. Some receivers may already search only during a cold start for the spreading sequences of satellites below the horizon. This test is obviously also very easy to circumvent by the designer of a signal simulator, which simply has to gradually attenuate signals as the simulated satellite's elevation reaches the horizon.

2.4.2 Power limits

With typical satellite altitudes of more than 20,000 km, the receiver–satellite distance, and therefore the best-case received signal strength, varies relatively little with elevation. It is guaranteed by the GPS specification to never exceed -150 dBW [14, 6.3.1]. A substantially stronger signal would indicate a manipulation. Power can be measured at different levels: (a) across the entire band, in form of the automatic gain control (AGC) signal, and (b) for a single satellite, in form of the correlation value reported by the prompt correlator in the code-tracking DLL. The across-the-band GPS L1 power level is largely dominated by thermal and receiver noise and therefore varies only little in normal operation. While a small amount of excess power beyond that, per satellite, can be explained by constructive multipath interference, anything stronger must be considered suspicious. On the other hand, there is no lower bound, as line-of-sight obstacles can always explain a lack of signal. Unlike a remote adversary, a local spoofing attacker should not find it difficult to adjust the power of the signal realistically, making this test less of a hurdle.

2.4.3 Doppler-shift verification

Many GNSS receivers track the phase of the received carrier signal, or more often that of a down-converted intermediate-frequency (IF) equivalent, after they have removed the ranging code, by implementing a Costas loop [15]. When such a loop has locked on, the input of its numerically controlled oscillator (NCO) is a function of the relative speed of both the transmitter and receiver antenna in an inertial coordinate system (Doppler shift, ± 10 kHz) as well as the frequency error of the local oscillator that is used to both down-convert and sample the incoming signal (typically a few parts per million). When the receiver tracks several satellites simultaneously, the frequency error of the local oscillator cancels out in the difference between the respective NCO inputs, and what remains (apart from tracking noise) is only the difference in the Doppler shifts between the satellites. A receiver can predict the Doppler shift of each satellite from the received ephemeris data and its own location and velocity, and compare these predictions with the observed Doppler-shift differences. The elimination of the local-oscillator error allows the application of tight tolerances in such checks, limited mainly by the uncertainty of the speeds involved and tracking noise. Such a test will require the designer of a simulator to accurately emulate the Doppler shift and will detect some comparatively simple simulators that do not.

Regular GNSS receivers will also estimate the Doppler shift in order to speed up initial signal acquisition, but may not apply any checks on the frequency once they are tracking a signal. They will try all reasonable Doppler shifts during a cold start. When connected to a signal simulator without accurate Doppler-shift generation, such receivers may take longer to acquire a signal but may otherwise not complain.

However, building a simulator that accurately reproduces Doppler shift is not that difficult. In a complex-number baseband representation of a quadrature-amplitude-modulated signal, Doppler shift Δf can be applied by multiplying the signal with $e^{2\pi i \Delta f t}$, thereby rotating the complex (or IQ) coordinate system with an angular velocity proportional to the Doppler shift. After several simulated individual satellite baseband signals have been frequency shifted this way, they can be added together before being fed into a single transmitter (with IQ input) that up-converts the signal to the carrier frequency. This is much cheaper than the individually tuned per-satellite transmitter claimed to be necessary in a Doppler-accurate simulator by [13].

2.4.4 Code–carrier phase comparison

The signal generators implemented in the satellites synthesize all aspects of the broadcast signal from a single atomic clock. As a result, the phases of all the emitted carriers and the pseudo-random-noise (PRN) code sequences and data signals modulated on top are strictly phase locked, i.e. there is a constant number of carrier periods per PRN chip and a constant number of PRN chips per data bit. Nevertheless, most receivers implement two independent tracking loops, a Costas loop for tracking the carrier and a PLL with early-late discriminator for tracking the

PRN spreading sequence. This is because most receivers first down-convert the microwave carrier band to an intermediate frequency of much less than 100 MHz. This frequency down-conversion introduces the frequency of the receiver's local oscillator as an additional variable and thereby destroys the fixed code–carrier phase relationship, making two tracking loops necessary for initial acquisition. Once both loops have locked on and the receiver switches from acquisition into tracking mode, many receivers use the feedback of the (less noisy) carrier-tracking loop to aid the PRN code tracking loop. [2, Ch. 5]

A signal simulator based on standard software-defined radio platforms (e.g., USRP) will digitally synthesize an IQ or IF signal that is then up-converted into the GNSS transmission band. Unless the synthesis of all the frequencies in this process is carefully phase locked and matched, the IF up-conversion process can easily break the fixed code–carrier phase relationship of a genuine signal. Regular receivers will not notice this during acquisition, and may not be disturbed by it much either during tracking, unless they do accurate phase accounting. Receivers that merely report a Doppler-shift frequency that crudely indicates the feedback signal in the carrier-tracking Costas loop are unlikely to help detect such deviations. What is needed instead is a register in each tracking loop that accurately integrates the frequency corrections that both tracking loops apply, in order to show the accumulated phase correction achieved (e.g., in metres). If this phase correction then starts to differ substantially between the carrier and code tracking loop, this would be a strong indication that the signal emerged from a simulator whose designer didn't worry too much about that phase relationship. Most normal GPS receivers do not accurately integrate the frequency correction onto a phase correction, however special carrier-based differential GPS receivers, used in some geodetic and robotic applications, may collect the raw data necessary to verify the code–carrier phase relationship.

2.4.5 Multi-band reception

A receiver that covers all the GNSS bands on which a satellite broadcasts (e.g. GPS L1 = 1.5754 GHz and L2 = 1.2276 GHz) can impose rather more substantial requirements on a signal simulator. In a genuine signal, the different carrier bands

- will be attenuated in nearly (but due to diffraction not exactly) the same way by line-of-sight obstacles;
- will show phase shifts caused by atmospheric diffraction, but remain phase locked.

A signal simulator might transmit only the signals in a single band (e.g., only GPS L1). If it broadcasts in multiple bands, it might lack the phase lock, phase shift, and close but imperfect power-level relationship typical of concurrently observed different carrier frequencies from the same satellite. Even if one of the carriers is modulated only with an unknown encrypted signal (e.g., Y on GPS L2), it can still be correlated against the same encrypted signal on any other carrier, in order to measure phase shift and compare attenuation.

2.4.6 Ephemeris data check

The orbital-position (ephemeris) data broadcast by each satellite should preferably be verified by comparing it with what is received at a secure reference receiver, or by verifying any cryptographic authenticity features included (digital signatures, message authentication codes, hash chains, etc.). GPS currently lacks the latter, but future systems might support cryptographic authenticity checks of ephemeris data.

Where neither of these options are feasible, a plausibility check against long-term invariants of the orbital data remains a possibility. Each satellite has a limited amount of fuel onboard, in order to change orbit, resulting in a maximum velocity change $||\Delta\mathbf{v}||$ achievable during its lifetime. This fuel can be used not only for station keeping, but also to reconfigure the orbital constellation, e.g. after satellite failures.[2] Likewise, satellite engines have limited thrust (especially ion engines), limiting the acceleration $||\Delta\mathbf{v}||/\Delta t$. If these limits and the rate of natural orbital perturbations are available, along with an algorithm that estimates a lower bound for the $||\Delta\mathbf{v}||$ needed to move the orbit of an satellite in a given time interval Δt from known past ephemeris data to the currently broadcast ones, these can be compared as a broad plausibility test.

However, the security gains achieved this way are limited: there appears to be no advantage to our local attacker from substantially deviating in the navigation data from the orbits of the satellites currently in space. False ephemeris data might be more useful in remote attacks, where the attacker wants to minimize the likelihood that the receiver reacquires the genuine signal, whereas we assume here the genuine signal to be easily suppressed.

2.4.7 Jump detection

Another commonly proposed type of spoofing detector looks for discontinuities in the received signals, e.g. the pseudo ranges or the resulting solutions for the location and local clock error, or bounds such changes with independent sensors (inertial navigation, odometer, dead reckoning, etc.). It is certainly prudent and practical to monitor the continuity of GNSS time against an independent, battery-backed local clock (see also Section 2.3.1). Such techniques also make sense to protect against remote attackers who start to spoof the signal after the receiver had already locked on to the genuine one. However, the applicability of such techniques against a local attacker seems rather limited, as the latter can replace the antenna with a signal generator while the receiver and alternative sensors are switched off. It also is not a practical instant check in situations where the GNSS receiver is only briefly switched on for an attestation operation, never running long enough to monitor the long-term continuity of satellite signals.

[2] The GPS satellites are rumoured to even be able to change the inclination of their orbits somewhat to achieve better polar coverage, should the need arise.

2.4.8 Quality metrics

Several quality metrics have been proposed in the literature for GNSS signals. If the quality of the received signal is substantially better than anything the receiver ever has seen with its real antenna attached, this might indicate the use of a signal generator. Examples of quality metrics include

- the residual error in the navigation solution (which solves an over-determined system of equations if more than four satellite are in sight);
- the deviation of the actual cross-correlation result from the ideal (e.g., triangular) autocorrelation function of the PRN signal.

2.5 Some other ideas

2.5.1 Individual transmitter characteristics

Signal analysis techniques have been developed that identify individual radio transmitters based on the influence that electronic component tolerances have on the exact shape of the emitted RF waveform. Parameters measured for transmitter fingerprinting include in particular

- carrier-frequency deviation;
- transients occurring when the carrier is switched on and off;
- amplitude and phase roll-off of the band-pass filters used to shape the output spectrum (which affect the shape of the eye pattern in digital modulation).

Normal GNSS signal generators are likely to use exactly the same mathematical function to synthesize the waveform for each satellite, adjusted only by obvious parameters such as Doppler shift, range phase shift, range attenuation, and spreading sequence. Real-world satellites may have additional other characteristics (hopefully within the tolerances allowed by the RF interface definition). However, carrier-frequency deviation is already carefully calibrated in GNSS signals, and as the signals are broadcast continuously, there is no opportunity to observe on/off transients. This leaves filter roll-off, which is difficult to measure directly given the very low signal-to-noise ratios typical of GNSS systems, especially where the spreading sequence is unknown to the receiver (e.g., GPS Y signal). It may show up, however, as satellite-individual and receiver-bandwidth dependent variations in the exact shape of the cross-correlation function.

2.5.2 Spectrum analysis

The RF input should normally see an expected minimum noise level not only within the transmission band (e.g., 20 MHz wide), but across the entire radio spectrum,

along with evidence of other, non-GNSS transmitters in adjacent bands. Substantial reduction of this out-of-spectrum noise level could indicate the use of a signal generator. This would require a more widely tuneable receiver to measure. An attacker who wants to fake this wider input spectrum would either have to use a substantially more wideband signal generator (more expensive, more power required), or would have to mix the synthesized GNSS spectral-band content with real background noise from an antenna (possibly with the GNSS band attenuated by a band-stop filter, or using spectrum frequency-shifted from a different band).

2.5.3 Extended search for GNSS signals

A regular GNSS receiver will lock onto a correlation peak with a particular spreading sequence as soon as one is found, or may search for a local maximum or the earliest peak among several nearby ones, in the interest of robust multi-path behaviour. A signal-authenticity verifying receiver could, in addition, continue to scan combinations of correlation-delay and Doppler-shift, and warn about the presence of more satellite signals than can be expected from the genuine transmitter constellation (e.g., the same spreading sequence at two range delays or Doppler shifts). This test is particularly useful if a local attacker mixes the simulated signal with background spectrum from an antenna, to evade the test outlined in the previous Section 2.5.2.

3 Comparison

The receiver technology required in order to implement the measures discussed in the preceding Section 2 differs substantially from method to method. Some require substantial extensions, or even alternative receiver architectures, compared to what is commonly implemented in existing civilian receivers. Commercial low-cost GPS chipsets receive only L1 C/A code. Most chips merely output time, location, and the identity and claimed azimuth and elevation of tracked satellites, using the very limited, but standardized, NMEA 0183 “sentences” ASCII format. Some GPS chips can also be switched into an additional, vendor-specific, binary communication protocol that gives access to additional data, such as the ephemeris, almanac and health information received from individual satellites. A very small number of GPS receiver chip sets provide even access to “raw” tracking data for each tracked satellite, such as pseudo range, Doppler shift, carrier–noise ratio, as well as internal receiver variables such as AGC gain setting, and local oscillator error from the navigation solution.

For many of the proposed methods, the only practical prototype implementation method involves a software-defined radio approach, where the 2–40 wide MHz GNSS band of interest is down-converted into an IQ baseband representation,

loaded block-by-block into RAM, and then all tracking and analysis algorithms are implemented in software [15].

Table 1 attempts to give an overview of the requirements and properties of each proposed method. The "Access" column describes at what level the measure needs to access the receiver's processing pipeline, and thereby gives an indication what existing GPS receiver chips could support such a measure: "RF" means that support has to be integrated in the RF front-end, "IQ" means that a software-defined receiver that receives downconverted IQ samples and then implements all further processing in software could implement the measure, "Raw" means that the proprietary binary protocols of some existing GPS receivers chips provide enough data, and "NMEA" means that the standard NMEA output of most existing GPS chips will suffice. The "Ref" column indicates whether communication with a separate, secure reference receiver station is required.

Table 1 Overview of the presented authenticity-verification methods

Method	Section	Access	Ref.	Extra requirements	Type
Conditional access	2.1.1	IQ		signal support, SIM	instant
Delayed release	2.1.2	IQ		signal support, NTP	instant
Permanently secret	2.1.3	IQ	Y	NTP	instant
Directional char.	2.2.1	Raw			cumulative
Impedance test	2.2.2	RF		TDR, etc.	instant
Time	2.3.1	NMEA		NTP, battery clock	instant
Navigation data	2.3.2	Raw	Y	NTP	both
Pseudo-ranges	2.3.3	Raw	Y	NTP	cumulative
Elevation limit	2.4.1	NMEA			instant
Power limits	2.4.2	Raw			instant
Doppler	2.4.3	Raw			instant
Code–carrier phase	2.4.4	IQ		or tracking-loop integrators	instant
Multiple bands	2.4.5	IQ		multiple down-converters	both
Ephemeris	2.4.6	Raw			instant
Jump	2.4.7	NMEA		battery-backed clock	cumulative
Quality metrics	2.4.8	IQ			both
Transmitter character.	2.5.1	IQ			both
Spectrum analysis	2.5.2	IQ		tuneable down-converter	both
Extended search	2.5.3	IQ			both

4 Conclusions

There clearly exist circumvention techniques for all the authenticity-verification methods outlined in this survey. The mechanisms available today for protecting GNSS signals against tampering by local attackers still can at best offer a level of security comparable to most other types of tamper-resistant hardware. They all fall well short of the ambition behind the Kerckhoffs' principle so popular in cryptology: detailed knowledge of the protection mechanisms used may still substantially

aid in their circumvention. Nevertheless, some of the presented mechanisms (e.g., secret spreading sequences, individual antenna characteristics) have the potential to prevent easy-to-use mass-market circumvention products. Others at least force the designer of a circumvention tool to add rather specialized functions, whose obvious purpose would be to circumvent these checks. The latter may help to enforce legal restrictions on their commercial availability. Some may be most useful as intrusion-detection tools that report suspicious signals for further investigation, rather than to automatically decide on their authenticity. In combination, they provide a formidable toolkit for managing the risk of local attackers on trusted GNSS receivers in many potential applications.

This work was supported by the European Commission under FP7 grant 228443 (TIGER project).

References

1. Parkinson BW, Spilker Jr. JJ (1996) Global Positioning System: theory and applications – Volume I. Progress in Astronautics and Aeronautics, Volume 163, American Institute of Aeronautics and Astronautics, Washington DC, ISBN 1-56347-106-X
2. Kaplan ED, Hegarty CJ (2006) Understanding GPS : principles and applications. 2nd edition, Artech House
3. Anderson RJ, Kuhn MG (1996) Tamper resistance — a cautionary note. In: The Second USENIX Workshop on Electronic Commerce Proceedings. USENIX Association, pp. 1–11
4. Weingart SH (2000) Physical security devices for computer subsystems: a survey of attacks and defenses. In: Cryptographic Hardware and Embedded Systems (CHES). LNCS 1965, Springer-Verlag, pp. 45–68
5. Anti-tamper physical security for electronic hardware. GORE `http://www.gore.com/en_xx/products/electronic/specialty/antitamper.html`
6. Joshi S (2008) Addressing the physical security of encryption keys. Maxim Engineering Journal, Vol. 62, pp. 7–11 `http://pdfserv.maxim-ic.com/en/ej/EJ62.pdf`
7. Anderson RJ (1998) On the security of digital tachographs. European Symposium on Research in Computer Security (ESORICS). LNCS 1485, Springer-Verlag, pp. 111–125
8. Humphreys TE, Ledvina BM, Psiaki ML, O'Hanlon BW, Kintner PM (2009, January) Assessing the spoofing threat. GPS World, p. 28–38
9. Garay JA, Staddon J, Wool A (2000) Long-lived broadcast encryption. Advances in Cryptology (CRYPTO). LNCS 1880, Springer-Verlag, p. 333–352
10. Kuhn MG (2004) An asymmetric security mechanism for navigation signals. In: 6th Information Hiding Workshop. LNCS 3200, Springer-Verlag, pp. 239–252
11. Scott L (2003) Anti-spoofing & authenticated signal architectures for civil navigation systems. ION GPS/GNSS, Institute of Navigation, pp. 1543–1552
12. Psiaki ML (2009) Spoofing detection for civilian GNSS signals via aiding from encrypted signals. ION GNSS
13. Wen H, Huang PY-R, Dyer J, Archinal J, Fagan J (2005). Countermeasures for GPS signal spoofing. ION GNSS
14. GPS Interface Control Document, ICD-GPS-200C, 2003-01-14
15. Borre K, Akos DM, Bertelsen N, Rinder P, Jensen SH (2007) A software-defined GPS and Galileo receiver. Birkhäuser