Metric Strand Spaces for Locale Authentication Protocols[*]

F. Javier Thayer, Vipin Swarup, and Joshua D. Guttman

The MITRE Corporation, USA
{jt,swarup,guttman}@mitre.org

Abstract. Location-dependent services are services that adapt their behavior based on the locations of mobile devices. For many applications, it is critical that location-dependent services use trustworthy device locations, namely locations that are both accurate and recent. These properties are captured by a security goal called *locale authentication* whereby an entity can authenticate the physical location of a device, even in the presence of malicious adversaries. In this paper, we present a systematic technique for verifying that location discovery protocols satisfy this security goal. We base our work on the strand space theory which provides a framework for determining which security goals a cryptographic protocol achieves. We extend this theory with a metric that captures the geometric properties of time and space. We use the extended theory to prove that several prominent location discovery protocols including GPS do not satisfy the locale authentication goal. We also analyze a location discovery protocol that does satisfy the goal under some reasonable assumptions.

1 Introduction

Location-dependent services have been well-studied and are widely expected to become an integral part of the pervasive computing infrastructure. These services adapt their behavior depending on the current locations of mobile devices. Device locations are provided by location discovery protocols. Security is a major concern in these systems and a wide variety of security goals have been studied in the literature, e.g., the privacy, freshness, and availability of location information. In this paper, we focus on the threat of malicious users deceiving location-dependent services by providing incorrect or stale location data. They may do this by subverting or bypassing the location discovery protocols.

Robustness against this threat is captured by the security goal of *locale authentication* whereby an entity can authenticate the physical location of a device, even in the presence of malicious adversaries. The entity can then issue a *locale certificate* that asserts that principal P has a physical presence at location (or within region) x at some time during a time interval t. A node that wishes to

[*] This work was supported by the MITRE-Sponsored Research Program.

M. Nishigaki et al. (Eds.): IFIPTM 2010, IFIP AICT 321, pp. 79–94, 2010.

use this attribute certificate must verify the certificate by checking its signature and determining whether the creator is trusted to have used a secure locale authentication protocol to authenticate the stated location.

Locale certificates are indeed feasible. Consider a person P who needs a certificate stating that she is at the physical address x at time t. P contacts a certificate granting authority which then dispatches a human agent to P's claimed location. The agent validates P's identity and secret key (which we assume P has previously obtained), verifies P's location at time t, and issues a document to P with this information and bearing the certificate authority's signature. Clearly this scheme is impractical since it requires a human to service each request, but it illustrates that the concept of a locale certificate is feasible.

A potentially better alternative is to provide each user with a tamper-proof GPS locator device that discovers and certifies the user's location. Such devices have been developed [4]. However, as we show in Section 7.2, any GPS device which relies exclusively on the commercial GPS external signal can be attacked and hence is inadequate for locale authentication in sensitive applications.

In this paper, we provide a precise definition of locale authentication and we present a systematic technique for proving whether or not location discovery protocols satisfy the locale authentication goal. Our work is based on the strand space theory which provides a framework for determining what security goals a cryptographic protocol achieves. Strand spaces are a special-purpose execution model for security protocols, based on a Lamport-style causal partial ordering. Behaviors of some principals, "regular principals," are assumed to follow the rules of the protocol, while others, adversarial principals, do whatever they like, constrained by cryptography and the secrets they possess.

The key notion is a possible global execution, or "bundle," namely a causally well-founded directed acyclic graph, in which the nodes are message transmissions or receptions. The nodes may lie on any number of regular or adversarial behaviors. Every message reception must be explained by some earlier transmission of the same message. Causal well-foundedness justifies a principle of induction for bundles, which is the basis of powerful and reusable proof techniques. Strand spaces were intended to capture a minimal view of protocol behavior, leading to conceptually spare and focused analysis techniques.

Secure location protocols combine cryptography with the physics of message transmission. The cryptographic operations authenticate the principals and preserve confidentiality, while the physics of message transmission constrain their possible locations. We enrich the strand space model by associating a space-time location with each node. The strands follow the world lines of principals. Some bundles are compatible with the physics of message transmission – e.g. the maximum message transmission speed – while others are not. An assertion true in every bundle compatible with the physics is a valid conclusion of a secure location protocol. We use this extended theory to prove that several prominent location discovery protocols including GPS do not satisfy the locale authentication goal. We also analyze a location discovery protocol that does satisfy the goal under some reasonable assumptions.

The remainder of this paper is organized as follows. In Section 3, we review related work. In Section 4, we summarize a broad range of security goals for location discovery protocols. We then focus on the goal of locale authentication and we describe our threat model and attack classes that compromise this goal. In Section 5, we present the metric strand space model. In Section 6, we consider four prominent location discovery protocols and we use the metric strand space model to examine whether they provide locale authentication under a specified threat model. Section 8 concludes the paper with a discussion of future directions for this novel line of work.

2 Some Examples

Consider a client device intended to participate in secure location protocols. It contains a private asymmetric key, and the manufacturer provides a certificate binding the public part of the same key pair to the identity of the purchaser. The device has some tamperproofing, to ensure that it is unlikely to have been subverted without effort on the part of the owner, and certainly not unbeknownst to the owner. The behavior of the device is as follows. When the owner specifies a number of location servers, the device interacts with them so that they can jointly provide evidence of the device's location.

The interaction between the device and the servers consists of two phases. First, a cryptographic protocol allows the device and server to agree on a pair of new shared secrets. This phase uses asymmetric cryptography and is thus expected to be slow. Second, the server transmits a random bitstring; when the device receives it, the device does a bitwise exclusive *or* with one shared secret, and replies with the result. This operation is very fast. The device also emits an estimate of the time t_δ elapsed between receipt and reply; the estimate is encrypted with the other shared secret. The server calculates the elapsed time t_ϵ between transmitting the challenge and receiving the *xor*ed response; it then certifies that the distance to the device is approximately $c \cdot (t_\epsilon - t_\delta)/2$, where c is the transmission speed in this medium. If the location of each server is known, and the device interacts with a suitable number of servers within a short period of time, then the location of the device on the surface of the earth or in three-dimensional space may be determined.

For the first phase of interaction, we use the Needham-Schroeder-Lowe protocol, taking the resulting shared secrets to be the two nonces. The full protocol is shown in Figure 1. Evidently, if it were possible for a third party to determine the nonces, then with judicious jamming, it may be possible for the third party to convince L that D is at the third party's location. For instance, if the protocol were based on the original, flawed Needham-Schroeder protocol instead of the version as fixed by Lowe, then an attack would be possible (Figures 2, 3). In this scenario, the owner has asked its device to determine its distance from P, a location server that turns out to act fraudulently; P can then surely convince an honest location server that D is where P is. By suitable choices of t_δ, probably any distance can be established.

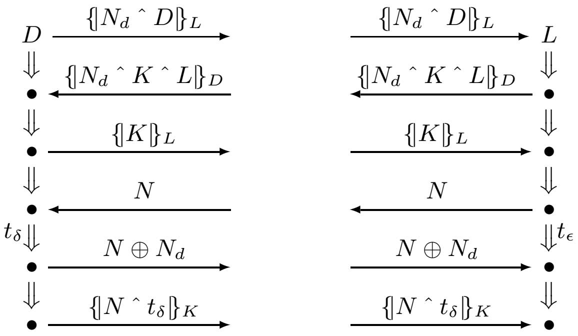

D
L
$\{\!|N_d \,\hat{}\, D|\!\}_L$
$\{\!|N_d \,\hat{}\, K \,\hat{}\, L|\!\}_D$
$\{\!|K|\!\}_L$
N
t_δ
t_ϵ
$N \oplus N_d$
$\{\!|N \,\hat{}\, t_\delta|\!\}_K$

Fig. 1. A Distance-determining Protocol

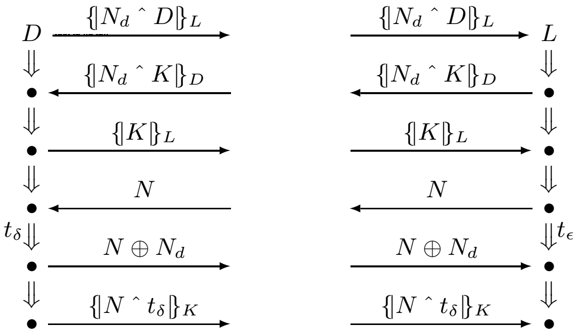

D
L
$\{\!|N_d \,\hat{}\, D|\!\}_L$
$\{\!|N_d \,\hat{}\, K|\!\}_D$
$\{\!|K|\!\}_L$
N
t_δ
t_ϵ
$N \oplus N_d$
$\{\!|N \,\hat{}\, t_\delta|\!\}_K$

Fig. 2. A Flawed Distance-determining Protocol

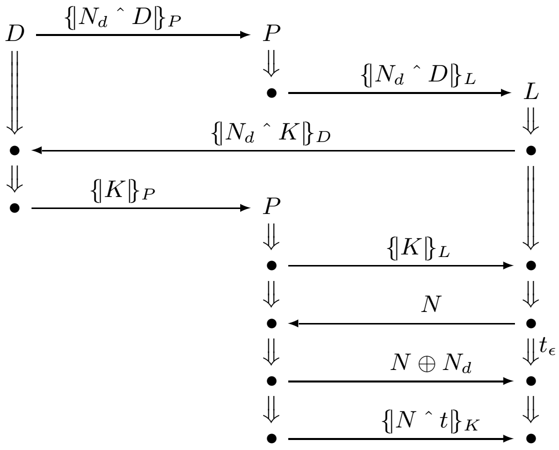

D
$\{\!|N_d \,\hat{}\, D|\!\}_P$
P
$\{\!|N_d \,\hat{}\, D|\!\}_L$
L
$\{\!|N_d \,\hat{}\, K|\!\}_D$
$\{\!|K|\!\}_P$
P
$\{\!|K|\!\}_L$
N
t_ϵ
$N \oplus N_d$
$\{\!|N \,\hat{}\, t|\!\}_K$

Fig. 3. An Attack on the Protocol of Figure 2

Thus, in protocols of this kind, there is a tight interconnection between the structure of the cryptographic protocol—the way that it uses cryptography to establish authentication and shared secrets—and its usefulness to determine location.

3 Related Work

A wide variety of location discovery protocols (e.g., [8]) have been proposed in the literature. Some protocols let a device compute its own location, some let a server compute the location of a device for the server's own use, and some let a server or device convince a third party of a device's location. The protocols may be passive where the load on location servers is independent of the number of protocol runs, or active where the server load is a function of the number of protocol runs (e.g., if the server has to respond to messages sent by devices).

Location discovery protocols compute the location of a device by using a distributed set of servers whose locations are known precisely. Distance bounding protocols [2] typically use distance or angle measurements of signal transmissions between the servers and the device, and they solve for the unknown location using triangulation (or multilateration) techniques. Distance measurements are made by either measuring the signal strength of a received signal or by measuring the arrival times of one or more signals. Signal strength measurements are converted to distance measurements by knowing the power of the transmitted signal and by modeling its expected strength at various locations [1]. Arrival times are converted to distance measurements by knowing the time of signal transmission and the propagation speed within the transmission medium [6]. Alternatively, if two signals with different propagation speeds are transmitted simultaneously, then the distance can be computed without knowing the actual transmission time of the signals [9]. Angle measurements are made by using special antenna configurations that measure the angle of arrival of a signal.

The most prominent location discovery system is the Global Positioning System (GPS). In this scheme, multiple synchronized satellites transmit periodic messages; a device that receives transmissions from several satellites can use the signals to triangulate its own position. Denning and MacDoran [4] have proposed building a locale authentication service using GPS signals. They suggest that a device can capture the signals it receives from multiple satellites and can package the received signals into a credential. A trusted server can verify that the signals have not been tampered with, and can use the signals to compute the device's position. As we prove in Section 7.2, such a scheme is not secure if the GPS geometry is public knowledge, e.g., as with commercial GPS.

Sastry et al [10] and Capkun et al [3] analyze protocols that can securely verify certain location claims. Their analyses are specific to their respective proposed protocols. Further, they only address the question of whether some arbitrary user is in a designated physical region. In contrast, in this paper we present a general, systematic model for verifying the security properties of any location discovery protocol. We apply this technique to several protocols, including a protocol that addresses the question of whether a specific user (i.e., a user with a specific secret key) is in a designated physical region. Meadows et al [7] present a qualitative framework for analyzing distance bounding protocols based on extending an authentication logic and is closest in spirit to our work.

4 Security for Location Discovery Protocols

Let N, M, P, and Q be principals (e.g., cryptographic keys, mobile nodes with identities, etc.), L be a physical location (or region), and t be a time interval. Some of N, M, P, and Q may possibly be the same principal. If a principal P has a physical presence at a location L (e.g., if P has access to a node at L), then we say that "P is at L" or that "P controls L".

The purpose of a location discovery protocol is to enable M to determine the location of N during t. However, location discovery protocols are subject to attacks by malicious penetrators. In this section, we examine a range of desirable security goals for location discovery protocols, and a set of attacks that may compromise those goals.

4.1 Security Goals

Desirable security goals for location discovery protocols include:

Locale Occupancy Authentication: If M determines that some principal was at L during t, then there was indeed a principal at L during t.

Locale Authentication: If M determines that N was at L during t, then N was indeed at L during t. Thus, a penetrator cannot make M believe that N is at a location other than N's true location.

Privacy: An unauthorized penetrator (e.g., an eavesdropping node) cannot obtain location information via a location discovery protocol. For instance, if M uses a location discovery protocol to determine that N is at L during t, then penetrator P should not be able to learn that N is at L during t.

Availability: An authorized principal M can determine N's location at any time. Thus, a penetrator cannot prevent M from determining N's location.

Nonrepudiation: Evidence that N was at L during t is irrefutable. With such evidence, M can prove to Q that N was at L during t.

Most of the literature on location discovery protocols addresses the functionality of the protocols, namely enabling M to determine the location of N during time interval t. Some techniques enable a node to determine its own location (e.g., [6], [9]), while others enable a node to determine the location of another node (e.g., by measuring signal strengths [1] or round-trip ping times). Some techniques address the privacy and availability goals, but most do not satisfy the goals of authentication or nonrepudiation.

4.2 Threat Model and Attacks

We distinguish between various threats based on the locations that are controlled by penetrators. Control of a location may enable a penetrator to monitor, delay, alter, delete, redirect, or replay any signals (message transmissions) that traverse that location or its immediate neighborhood. However, in this paper, we assume that a penetrator cannot block broadcast transmissions between two neighboring nodes that the penetrator does not control. Note that some penetrators may only control a small number of locations (e.g., a soldier in enemy territory). Some penetrators may control an entire physical region (e.g., an enemy state that controls some territory). Some penetrators ("insiders") may control some nodes used by location discovery protocols, e.g., trusted location servers.

Location discovery protocols are subject to a wide range of attacks. For instance, a penetrator may deny that he was at location L during time t, a penetrator may prevent M from determining N's location, a penetrator may monitor the location discovery protocol to learn the location of a user, etc. In this paper, however, we focus only on attacks on authentication. There are two broad types of attacks.

Location integrity attacks may cause M to believe that N is at location L, even though N does not have a physical presence at L. Two special cases of this attack are: (a) *deception attacks* where a penetrator P causes M to believe that P is at location L, even though that is not the case; and (b) *positioning attacks* where a penetrator P causes M to believe that M is at location L, even though that is not the case. Note that, depending on the attack, the penetrator may or not be able to select the location L.

Masquerade attacks by a penetrator P at L may cause M to believe that the penetrator is N at L. Note that a penetrator P at L may launch a combination masquerade and location integrity attack to cause M to believe that N is at L'.

5 Metric Strand Spaces

We now extend the formalism of strand spaces [11] (see Appendix A for a summary of the strand space theory). The extended theory will enable us to prove whether a location discovery protocol is secure against location integrity and masquerade attacks. In our intended application, a strand will model a sequence of events in a process or in a processor.

We assume that all nodes are subject to physical constraints dictated by the laws of nature as we understand them. We use two laws in this paper. First, a message must be sent before it is received. Second, messages cannot travel faster than the signal speed in the medium of transmission (e.g., the speed of sound in water, or the speed of light). We also assume an idealized model of message transmission; specifically, we assume that broadcast messages are transmitted by a spherical wavefront.

A *partial pseudometric* on a set X is a partial function d with non-negative real values on $X \times X$ such that $d(x,x) = 0$, $d(x,y) = d(y,x)$ when either side is defined, and the triangle inequality

$$d(x, y) \leq d(x, z) + d(z, y)$$

holds when $d(x, y)$, $d(x, z)$, $d(z, y)$ are all defined.

A *time elapse function* on a partially ordered set $(X, \preceq)$ is a partial function e with non-negative real values defined on pairs x, y where $x \preceq y$ such that $\mathrm{e}(x, x) = 0$ and with the additivity property:

$$\mathrm{e}(x, y) = \mathrm{e}(x, z) + \mathrm{e}(z, y) \tag{1}$$

whenever $x \preceq z \preceq y$.

Recall that for nodes m, m' in a single strand of a strand space, $m \Rightarrow^* m'$ means there is a sequence of intermediate nodes $m \Rightarrow m_1 \Rightarrow m_2 \Rightarrow \cdots \Rightarrow m'$, all on the same strand.

Definition 1. *A* geometric bundle *is a tuple* $(\mathcal{B}, \mathrm{dist}, \mathrm{elapse}, c)$ *where* $\mathcal{B}$ *is a bundle,* dist *is a partial pseudometric on* $\mathrm{nodes}(\mathcal{B})$*,* elapse *is a time elapse function on pairs* $m, n \in \mathrm{nodes}(\mathcal{B})$ *such that* $m \Rightarrow^* n$*, and* c *is a positive real.*

Notice that it is possible for different nodes m, n to be within 0 distance of each other. For instance, if the strand s is associated to a static object, then all nodes on s are within 0 distance of each other. However, we can obtain a metric space $\tilde{X}$ from $X = \mathrm{nodes}(\mathcal{B})$, by identifying points m, n which are at 0 distance from each other. The space $\tilde{X}$ and the quotient map $x \mapsto \tilde{x}$ completely determine X. We will refer to $\tilde{X}$ as the geometry of the bundle, c as the message propagation speed, and elapse as the elapsed time between successive events on a strand. We need to relate the metric and the elapsed time function:

Axiom 1. *Suppose* $m_1 \rightarrow n_1 \Rightarrow^* m_2 \rightarrow \cdots \Rightarrow^* m_k \rightarrow n_k$ *is a path in the bundle such that* m_1, n_k *are on the same strand and* $\mathrm{dist}(m_k, n_k)$ *are all defined. Then*

$$\mathrm{elapse}(m_1, n_k) = \frac{1}{c} \sum_{i=1}^{k} \mathrm{dist}(m_i, n_i) + \sum_{\ell=1}^{k-1} \mathrm{elapse}(n_\ell, m_{\ell+1}). \tag{2}$$

In this paper, elapse will be interpreted as elapsed time with respect to a global clock. A *global time* on a bundle is a function $\mathrm{T} : \mathrm{nodes}(\mathcal{B}) \rightarrow \mathbb{R}$ such that

$$\mathrm{elapse}(m, n) = \mathrm{T}(m) - \mathrm{T}(n).$$

Given a geometric bundle with global time, we will use terms such as propagation time, propagation distance, etc.

If strands may be associated with mobile objects, we need to relate the metric and the elapsed time function for nodes on a single strand: $\mathrm{dist}(m, n) \leq \mathrm{v} \times \mathrm{elapse}(m, n)$ where $m \Rightarrow^* n$ and where v is a positive real that represents the maximum speed of any object. However, for simplicity, we assume in this paper that all nodes on a strand have the same location (the results in this paper can be extended easily to deal with mobile objects by using the above axiom). This corresponds to strands being associated with static objects and is captured by the following definition:

Definition 2. *A* static geometric bundle *is a geometric bundle* $(\mathcal{B}, \text{dist}, \text{elapse}, c)$ *where for all pairs* $m, n \in \text{nodes}(\mathcal{B})$ *such that* $m \Rightarrow^* n$*,* $dist(m, n) = 0$*.*

6 Security Analysis of Location Protocols

We now give several illustrative examples of protocols, with precise locale authentication claims and proofs.

6.1 Protocol 1

This first protocol is ancient. It is typically used in the inverse direction to empirically determine the medium propagation speed c from the known distance.

The initiator A sends a nonce N_a and expects return N_a. Responder returns N_a. There are thus two kinds of strands in this protocol:

1. Initiator strands $s = \langle m_1, m_2 \rangle \in \text{Init}[N_a]$ with trace $\langle +N_a, -N_a \rangle$.
2. Responder strands $t = \langle n_1, n_2 \rangle \in \text{Resp}[N_a]$ with trace $\langle -N_a, +N_a \rangle$.

Proposition 1. *Suppose* $\mathcal{C}$ *is a static geometric bundle. If* $\mathcal{C}$ *contains a strand* $s = \langle m_1, m_2 \rangle \in \text{Init}[N_a]$ *of height* 2 *and* N_a *uniquely originates on* s*, then* $\mathcal{C}$ *contains a strand whose distance from the nodes of* s *is at most* $c/2{\times}\big(\text{elapse}(m_1, m_2)\big)$*.*

Proof. Let $s = (m_1, m_2)$. Since N_a uniquely originates on m_1, there is a path $p_1 \to q_1 \Rightarrow^* p_2 \to \cdots \Rightarrow^* p_k \to q_k$ from m_1 to m_2 (i.e., $p_1 = m_1$ and $q_k = m_2$). Now:

$$\begin{aligned}\text{elapse}(m_1, m_2) &\geq \frac{1}{c}\sum_{i=1}^{k} \text{dist}(p_i, q_i) \geq \frac{1}{c}(\text{dist}(p_1, p_2) + \text{dist}(p_2, q_k))\\ &= 2 \times \frac{1}{c} \times \text{dist}(m_1, p_2) \quad (\text{since } p_1 = m_1 \text{ and } q_k = m_2)\end{aligned}$$

Thus, $\text{dist}(m_1, p_2) \leq c/2 \times \text{elapse}(m_1, m_2)$. ■

Since the strand whose existence is claimed may be a penetrator strand, this protocol gives no information about who is within the specified distance.

6.2 Protocol 2

The purpose of this protocol is to provide A with a number $r > 0$ and the following guarantee to A: B is within distance r from A. A generates a nonce value which it signs and sends to B. B receives N_a and determines a time delay T to allow for processing time. B signs $N_a\,\hat{}\,T$ and after an elapsed time of T sends $N_a\,\hat{}\,T$ with its signature to A.

Schematically, the protocol can be described thus:

$$\begin{tikzcd}
\text{A} & \text{B} \\
\bullet \arrow[r, "\{\!|N_a|\!\}_{K_A^{-1}}"] \arrow[d, Rightarrow] & \bullet \arrow[d, Rightarrow] \\
\bullet & \bullet \arrow[l, "\{\!|N_a\,\hat{}\,T|\!\}_{K_B^{-1}}"']
\end{tikzcd}$$

Note that we are modeling a message accompanied by a hash as an encryption with a private key. After running this protocol A should have B's reply. Let T_{elapsed} be the time interval A observes between the transmission of the nonce N and receipt of the response from B. A then computes the value:

$$r = \frac{1}{2} \times \big(T_{\text{elapsed}} - T\big) \times c$$

Of course an attacker could still be present, but the claim is that A has received a guarantee that B, an honest participant, is within a ball of radius r of A.

There are two kinds of strands in this protocol:

1. Initiator strands $s = \langle m_1, m_2\rangle \in \text{Init}[B, N_a, T]$ with trace
$$\langle +N_a, -\{\!|N_a \mathbin{\hat{}} T|\!\}_{K_B^{-1}}\rangle.$$
2. Responder strands $t = \langle n_1, n_2\rangle \in \text{Resp}[B, N_a, T]$ with trace
$$\langle -N_a, +\{\!|N_a \mathbin{\hat{}} T|\!\}_{K_B^{-1}}\rangle$$
for which $\text{elapse}(n_1, n_2) \geq T$.

We need to assume the hash-key mapping $A \mapsto K_A^{-1}$ is injective.

Proposition 2. *Suppose $\mathcal{C}$ is a static geometric bundle. Suppose also that the set of penetrator keys is empty. If $\mathcal{C}$ contains a strand s in* Init$[B, N_a, T]$ *of height* 2 *and N_a uniquely originates on s, then $\mathcal{C}$ contains a strand t in* Resp$[B, N_a, T]$ *of height* 2*. Moreover the distance from the nodes of s to the nodes of t is at most* $1/2 \times \big(\text{elapse}(m_1, m_2) - T\big) \times c$.

Proof. Let $s = \langle m_1, m_2\rangle$. The term $\{\!|N_a \mathbin{\hat{}} T|\!\}_{K_B^{-1}} = \text{term}(m_2)$ originates on at least one strand t. t is regular, since the set of penetrator keys is empty. Thus t must be the second node n_2 of a regular strand in Resp$[B, N_a, T]$. In particular there is a path $q = \langle q_0, q_1, \ldots, q_{\ell(q)}\rangle$ from n_2 to m_2. Since N_a is uniquely originating at s, there is a path $p = \langle p_0, p_1, \ldots, p_{\ell(p)}\rangle$ from the first node m_1 of s to the first node n_1 of t. The paths p, q may traverse intermediate nodes on various strands. However, by the definition of geometric strand space and Axiom 1:

$$\begin{aligned}
&\text{elapse}(m_1,m_2) \geq \\
&\quad \geq \frac{1}{c} \sum_{k=0}^{\ell(p)-1} \text{dist}(p_k, p_{k+1}) + \text{elapse}(n_1, n_2) + \frac{1}{c} \sum_{k=0}^{\ell(q)-1} \text{dist}(q_k, q_{k+1}) \\
&\quad \geq \frac{1}{c}\,\big(\text{dist}(m_1, n_1) + \text{dist}(m_2, n_2)\big) + T = 2 \times \frac{1}{c} \times \text{dist}(m_1, n_1) + T
\end{aligned}$$

Thus, $\text{dist}(m_1, n_1) \leq 1/2 \times \big(\text{elapse}(m_1, m_2) - T\big) \times c$. ■

6.3 Clock Distortion

The preceding protocol has two related problems:

1. The condition that the reported delay time T be less than the elapsed time for the initiator is too severe;
2. The requirement that principals have exact clocks is unrealistic.

We briefly indicate a simple refinement of the model that can deal with bounded time drift and distortion. The main idea here is that the time variation is bounded:

Definition 3. *A time elapse function τ_s on a strand s is α bounded iff*

$$\mathrm{elapse}(m,n) \leq \alpha\ \tau(m,n).$$

Having these distortion bounds allows us to introduce locale certificates with specified tolerances. We leave the details to future work.

7 Noninteractive Schemes

The previous locale schemes were based on the pessimistic assumption that the attacker may actually block all messages. We now consider schemes in which this capability is weakened or eliminated altogether. The impossibility of blocking messages is formulated as follows:

Definition 4. *A bundle $\mathcal{C}$ in a geometric strand space with global time* T *is* radial *iff for every positive node m and every strand s there is a negative node n on s such that $m \rightarrow n$ and* $\mathrm{T}(n) = \mathrm{T}(m) + \frac{1}{c}\ \mathrm{dist}(m,n)$

The property expressed in this definition implies that messages sent are always received by all principals and in particular, that messages cannot be destroyed by the attacker. Questions of system failure are not an issue here since there is no intermediate hardware between sender and receiver to required for message delivery. A weaker property which allows for failure conditions for strands is also possible, but our goal here is to consider questions of authenticity not of liveness.

7.1 Protocol 3

Consider the following protocol: A generates a nonce value N, attaches a signature and emits the signed term $\{\!|N|\!\}_{K_A^{-1}}$. The value $\{\!|N|\!\}_{K_A^{-1}}$ is received by a set of location servers B_i. We assume the location servers B_i have access to universal time T. B_i notes the time $T_i = \mathrm{T}(m_i)$ it first receives the signal $\{\!|N|\!\}_{K_A^{-1}}$. We note that B_i may receive the signal more than once since an attacker may attempt to confuse the location servers by delay and retransmission, The servers B_i send the values $\{\!|\{\!|N|\!\}_{K_A^{-1}} \,\hat{}\, T_i|\!\}_{K_i^{-1}}$ to a central server, who creates a certificate with information on A's location.

$$\begin{tikzcd}
\mathrm{A} & \mathrm{B}_i & & \mathrm{S} \\
n_A \arrow[r, "\{\!|N|\!\}_{K_A^{-1}}"] & m_i \arrow[d, Rightarrow] & & \\
& \bullet \arrow[rr, "\{\!|\{\!|N|\!\}_{K_A^{-1}} \,\hat{}\, T_i|\!\}_{K_i^{-1}}"] & & \bullet
\end{tikzcd}$$

Proposition 3. *The following relations hold:*

$$\frac{1}{c}\Bigl(\mathrm{dist}(n_A, m_i) - \mathrm{dist}(n_A, m_j)\Bigr) = T_i - T_j. \tag{3}$$

Proof. Suppose A generates the message at time $T_A = \mathrm{T}(n_A)$ then by the definition of global time, for every B_i we have the identity:

$$T_i - T_A = \frac{1}{c}\,\mathrm{dist}(n_A, m_i),$$

from which formula (3) follows immediately by subtraction.

Corollary 4. *Suppose the underlying geometry is $\mathbb{R}^3$ with Euclidean distance. Then the distance equation for each pair of servers constrains n_A to be on a hyperboloid.*

Proof. This follows from the fact that the locus of points z whose distances from points a_0, a_1 satisfy "$\mathrm{dist}(z, a_0) - \mathrm{dist}(z, a_1) = \text{constant}$" is a hyperboloid. ■

This scheme is impractical since it requires that all servers respond to each request for locale authentication.

7.2 Protocol 4

We now describe a passive scheme much like GPS. The scheme is passive in the sense that the location servers just transmit signals instead of responding to clients. The protocol we describe is an idealized form of the actual protocol because we assume transmissions are continuous. For each time value t the server A_i generates a nonce N_{t,A_i} and at time t transmits $\{\!|N_{t,A_i}|\!\}_{K_{A_i}^{-1}}$. The client collects all signals received at a single time instant and packages the messages into a vector $\langle \{\!|N_{t,A_i}|\!\}_{K_{A_i}^{-1}}, \ldots, \{\!|N_{t,A_i}|\!\}_{K_{A_i}^{-1}} \rangle$. This vector is used as a locale certificate. It would appear to constrain the client to be on an intersection of balls whose centers are the locations of the servers.

Unfortunately, this scheme does not appear to be much better than saying "I am here", because nothing enforces the client to package the messages actually received at a single instant, even if the servers encrypt the signals. Encryption of the nonces adds no security to this protocol, since an attacker may determine beforehand which value of t to use for each server.

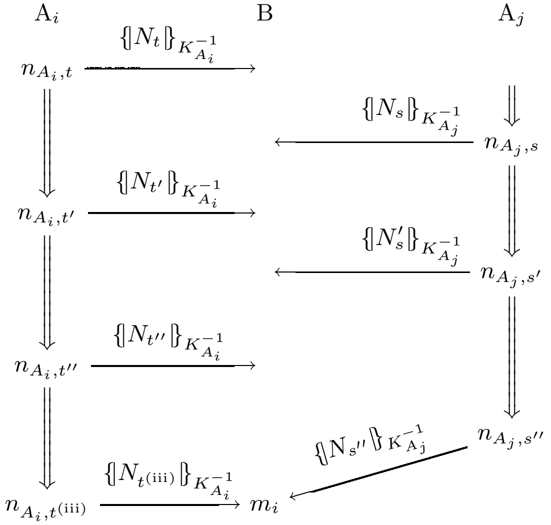
To analyze this protocol with encrypted nonces, we associate to each server A_i the function $\mu_i : t \mapsto \{N_{t,A_i}\}_{K_{A_i}}$. For each time instant t, the value of this function is the encrypted value of the nonce transmitted at time t by the server A_i. Let us consider a fixed client B with location x. Since we are assuming the underlying geometry is radial, any signal transmitted by the server is received by the client after propagation delay. Denoting the signal propagation delay to x from the location of A_i by $\tau_{A_i,x}$, we have

$$\tau_{A_i,x} = \frac{1}{c}\,\mathrm{dist}(A_i, x) \tag{4}$$

Thus the client sees a time translate of the server generated function μ_i

$$\phi_{A_i,x}(t) = \mu_i(t - \tau_{A_i,x}). \tag{5}$$

The goal of this protocol, ostensibly, is to provide a principal at x the data to assemble a locale certificate that binds that principal to location x at time t. The client is expected to package the sequence of values corresponding to messages received at time t, into a vector $\langle \phi_{A_1,x}(t), \ldots, \phi_{A_n,x}(t)\rangle$. which he can then use as a certificate. However, a malicious client may package some other sequence of signals received at different times $t_1, \ldots, t_n$ into a bogus certificate $\langle \phi_{A_1,x}(t_1), \ldots, \phi_{A_n,x}(t_n)\rangle$.


A_i
B
A_j
$\{\!|N_t|\!\}_{K^{-1}_{A_i}}$
$n_{A_i,t}$
$\{\!|N_s|\!\}_{K^{-1}_{A_j}}$
$n_{A_j,s}$
$\{\!|N_{t'}|\!\}_{K^{-1}_{A_i}}$
$n_{A_i,t'}$
$\{\!|N'_s|\!\}_{K^{-1}_{A_j}}$
$n_{A_j,s'}$
$\{\!|N_{t''}|\!\}_{K^{-1}_{A_i}}$
$n_{A_i,t''}$
$\{\!|\mathrm{N}_{s''}|\!\}_{\mathrm{K}^{-1}_{\mathrm{A_j}}}$
$n_{A_j,s''}$
$\{\!|N_{t^{(\mathrm{iii})}}|\!\}_{K^{-1}_{A_i}}$
$n_{A_i,t^{(\mathrm{iii})}}$
m_i

In the following proposition we use the expression "complete knowledge of the underlying geometry" whose meaning is explained in the context of the proof.

Proposition 5. *Under free encryption, if B has complete knowledge of the underlying geometry, of its own location and that of all the servers, then B can falsify any location.*

Proof. B is at location x. Let y be some arbitrary location. Since the locations of the servers are known to B, the distances $\mathrm{dist}(A_i, y)$ are known to B. It follows that the propagation times

$$\tau_{A_i,y} = \frac{1}{c}\,\mathrm{dist}(A_i, y)$$

from location of A_i to y are also known to B. Note

$$\begin{aligned} \phi_{A_i,y}(t) &= \mu_i(t - \tau_{A_i,y}) \\ &= \mu_i((t + \tau_{A_i,x} - \tau_{A_i,y}) - \tau_{A_i,x}) \\ &= \phi_{A_i,x}(t + \tau_{A_i,x} - \tau_{A_i,y}). \end{aligned} \tag{6}$$

Every term in this last expression is accessible to B. Thus B can claim the vector

$$\langle \phi_{A_1,x}(t + \tau_{A_1,x} - \tau_{A_1,y}), \ldots, \phi_{A_n,x}(t + \tau_{A_n,x} - \tau_{A_n,y})\rangle.$$

as a time t location certificate, thus asserting to be at location y. ■

We leave open the more general question of whether it is possible to have a passive location discovery protocol that satisfies the locale authentication goal.

8 Conclusion

We identified the security goal of locale authentication and provided a systematic technique for proving that location protocols satisfy that goal. The technique is based on extending the well-developed strand space theory with a metric that captures the geometric properties of time and space. We used the metric strand space theory to prove that several prominent location discovery protocols including GPS do not satisfy the locale authentication goal. We also analyzed a location discovery protocol that does satisfy the goal under some reasonable assumptions.

There are occasions in which certificates for physical parameters other than location would be desirable. For instance, a regulatory agency may require certification that certain parameters (level of contaminants, radiation, temperature) in a production facility are within legal tolerances. In principle, this can be done in much the same way by a certification authority by dispatching a human agent to the site. This can also be done by a tamper-proof gauge which issues an electronic certificate or prints out a paper one. However, there is no obvious way to exploit some physical resource such as geometry so that these certificates can be issued remotely without tamper-proof gadgetry. At this point the remote construction of such certificates remains an open problem. Another interesting question is whether it is possible to have a passive location discovery protocol that satisfies the locale authentication goal.

References

1. Bahl, P., Padmanabhan, V.N.: Radar: An in-building RF-based user location and tracking system. In: Proceedings of IEEE INFOCOM, vol. 2, pp. 775–784 (2000)
2. Brands, S., Chaum, D.: Distance-bounding protocols. In: Helleseth, T. (ed.) EUROCRYPT 1993. LNCS, vol. 765, pp. 344–359. Springer, Heidelberg (1994)
3. Capkun, S., Buttyan, L., Hubaux, J.P.: SECTOR: Secure tracking of node encounters in multi-hop wireless networks. In: Proceedings of the ACM Workshop on Security of Ad hoc and Sensor Networks, SASN (2003)
4. Denning, D.E., MacDoran, P.F.: Location-based authentication: Grounding cyberspace for better security. Computer Fraud & Security (February 1996)
5. Guttman, J.D., Thayer Fábrega, F.J.: Authentication tests and the structure of bundles. Theoretical Computer Science 283(2), 333–380 (2002)
6. Hofmann-Wellenhof, B., Lichtenegger, H., Collins, J.: The Global Positioning System: Theory and Practice. Springer, Wien (20014b)
7. Meadows, C., Poovendran, R., Pavlovic, D., Chang, L., Syverson, P.: Distance bounding protocols: Authentication logic analysis and collusion attacks. In: Advances in Information Security, vol. 30, pp. 279–298. Springer, Heidelberg (2007)
8. Poovendran, R., Wang, C., Roy, S. (eds.): Secure Localization and Time Synchronization for Wireless Sensor and Ad Hoc Networks. Advances in Information Security, vol. 30. Springer, Heidelberg (2007)
9. Priyantha, N.B., Miu, A.K.L., Balakrishnan, H., Teller, S.J.: The cricket compass for context-aware mobile applications. In: MOBICOM, pp. 1–14 (2001)
10. Sastry, N., Shankar, U., Wagner, D.: Secure verification of location claims. In: Proceedings of the ACM Workshop on Wireless Security, WiSe (2003)
11. Thayer Fábrega, F.J., Herzog, J.C., Guttman, J.D.: Strand spaces: Proving security protocols correct. Journal of Computer Security 7(2/3), 191–230 (1999)

A Strand Space Definitions

This appendix, derived from [11,5], defines the basic strand space notions.

A.1 Strands, Strand Spaces, and Origination

Consider a set A, the elements of which are the possible messages that can be exchanged between principals in a protocol. We will refer to the elements of A as *terms*. In a protocol, principals can either send or receive terms. We represent transmission of a term as the occurrence of that term with positive sign, and reception of a term as its occurrence with negative sign.

Definition 5. *A* signed term *is a pair* $\langle \sigma, a \rangle$ *with* $a \in \mathsf{A}$ *and* σ *one of the symbols* $+, -$. *We will write a signed term as* $+t$ *or* $-t$. $(\pm\mathsf{A})^*$ *is the set of finite sequences of signed terms. We will denote a typical element of* $(\pm\mathsf{A})^*$ *by* $\langle\, \langle \sigma_1, a_1 \rangle, \ldots, \langle \sigma_n, a_n \rangle\, \rangle$.

A strand space *over* A *is a set* Σ *with a trace mapping* $\mathrm{tr} : \Sigma \to (\pm\mathsf{A})^*$.

By abuse of language, we will still treat signed terms as ordinary terms. For instance, we shall refer to subterms of signed terms. We will usually represent a strand space by its underlying set of strands Σ.

Definition 6. *Fix a strand space Σ.*

1. *A* node *is a pair $\langle s, i\rangle$, with $s \in \Sigma$ and i an integer satisfying $1 \leq i \leq length(tr(s))$. The set of nodes is denoted by $\mathcal{N}$. We will say the node $\langle s, i\rangle$ belongs to the strand s. Clearly, every node belongs to a unique strand.*
2. *If $n = \langle s, i\rangle \in \mathcal{N}$ then $index(n) = i$ and $strand(n) = s$. Define $term(n)$ to be $(tr(s))_i$, i.e. the ith signed term in the trace of s. Similarly, $uns\_term(n)$ is $((tr(s))_i)_2$, i.e. the unsigned part of the ith signed term in the trace of s.*
3. *There is an edge $n_1 \rightarrow n_2$ if and only if $term(n_1) = +a$ and $term(n_2) = -a$ for some $a \in \mathsf{A}$. Intuitively, the edge means that node n_1 sends the message a, which is received by n_2, recording a potential causal link between those strands.*
4. *When $n_1 = \langle s, i\rangle$ and $n_2 = \langle s, i + 1\rangle$ are members of $\mathcal{N}$, there is an edge $n_1 \Rightarrow n_2$. Intuitively, the edge expresses that n_1 is an immediate causal predecessor of n_2 on the strand s. We write $n' \Rightarrow^+ n$ to mean that n' precedes n (not necessarily immediately) on the same strand.*
5. *An unsigned term t* occurs in *$n \in \mathcal{N}$ iff $t \sqsubset term(n)$.*
6. *Suppose I is a set of unsigned terms. The node $n \in \mathcal{N}$ is an* entry point *for I iff $term(n) = +t$ for some $t \in I$, and whenever $n' \Rightarrow^+ n$, $term(n') \notin I$.*
7. *An unsigned term t* originates *on $n \in \mathcal{N}$ iff n is an entry point for the set $I = \{t' : t \sqsubset t'\}$.*
8. *An unsigned term t is* uniquely originating in *a set of nodes $S \subset \mathcal{N}$ iff there is a unique $n \in S$ such that t originates on n.*
9. *An unsigned term t is* non-originating in *a set of nodes $S \subset \mathcal{N}$ iff there is no $n \in S$ such that t originates on n.*

If a term t originates uniquely in a suitable set of nodes, then it can play the role of a nonce or session key, assuming that everything that the penetrator does in some scenario is in that set of nodes.

$\mathcal{N}$ together with both sets of edges $n_1 \rightarrow n_2$ and $n_1 \Rightarrow n_2$ is a directed graph $\langle \mathcal{N}, (\rightarrow \cup \Rightarrow)\rangle$.

A *bundle* is a finite subgraph of $\langle \mathcal{N}, (\rightarrow \cup \Rightarrow)\rangle$, for which we can regard the edges as expressing the causal dependencies of the nodes.

Definition 7. *Suppose $\rightarrow_\mathcal{C} \subset \rightarrow$; suppose $\Rightarrow_\mathcal{C} \subset \Rightarrow$; and suppose $\mathcal{C} = \langle \mathcal{N}_\mathcal{C}, (\rightarrow_\mathcal{C} \cup \Rightarrow_\mathcal{C})\rangle$ is a subgraph of $\langle \mathcal{N}, (\rightarrow \cup \Rightarrow)\rangle$. $\mathcal{C}$ is a bundle if:*

1. *$\mathcal{N}_\mathcal{C}$ and $\rightarrow_\mathcal{C} \cup \Rightarrow_\mathcal{C}$ are finite.*
2. *If $n_2 \in \mathcal{N}_\mathcal{C}$ and $term(n_2)$ is negative, then there is a unique n_1 such that $n_1 \rightarrow_\mathcal{C} n_2$.*
3. *If $n_2 \in \mathcal{N}_\mathcal{C}$ and $n_1 \Rightarrow n_2$ then $n_1 \Rightarrow_\mathcal{C} n_2$.*
4. *$\mathcal{C}$ is acyclic.*

In conditions 2 and 3, it follows that $n_1 \in \mathcal{N}_\mathcal{C}$, because $\mathcal{C}$ is a graph.

Definition 8. *A node n is in a bundle $\mathcal{C} = \langle \mathcal{N}_\mathcal{C}, \rightarrow_\mathcal{C} \cup \Rightarrow_\mathcal{C}\rangle$, written $n \in \mathcal{C}$, if $n \in \mathcal{N}_\mathcal{C}$; a strand s is in $\mathcal{C}$ if all of its nodes are in $\mathcal{N}_\mathcal{C}$.*

*If $\mathcal{C}$ is a bundle, then the $\mathcal{C}$-*height *of a strand s is the largest i such that $\langle s, i\rangle \in \mathcal{C}$. $\mathcal{C}$-*trace(s) *$= \langle tr(s)(1), \ldots, tr(s)(m)\rangle$, where $m = \mathcal{C}$-height(s).*

We say that $s \in \mathcal{C}$ if the $\mathcal{C}$-height of s equals length(s).