On the Establishment of Distinct Identities in Overlay Networks

Rida A. Bazzi
Computer Science and Engineering
Arizona State University
Tempe, Arizona
bazzi@asu.edu

Goran Konjevod*
Computer Science and Engineering
Arizona State University
Tempe, Arizona
goran@asu.edu

ABSTRACT

We study ways to restrict or prevent the damage that can be caused in a peer-to-peer network by corrupt entities creating multiple pseudonyms. We show that it is possible to *remotely* issue certificates that can be used to test the distinctness of identities. To our knowledge, this is the first work that shows that remote anonymous certification of identity is possible under adversarial conditions. Our certification protocols are based on geometric techniques that establish location information in a fault-tolerant and distributed fashion. They do not rely on a centralized certifying authority or infrastructure that has direct knowledge of entities in the system, and work in Euclidean or spherical geometry of arbitrary dimension. Our protocols tolerate corrupt entities, including corrupt certifiers as well as collusion by certification applicants and certifiers. We consider both broadcast and point-to-point message passing models.

Categories and Subject Descriptors

C.2.4 [**Computer-communication networks**]: Distributed systems; H.3.4 [**Information storage and retrieval**]: Distributed systems; C.2.2 [**Network protocols**]: Applications; K.6.5 [**Computing milieux**]: Manatement of computing and information systems—*security and protection*

General Terms

algorithms, security, theory, verification

Keywords

sybil attack, identity verification, security, overlay networks, peer-to-peer systems, fault-tolerance, distance geometry

*Supported in part by the NSF Grant CCR-0209138

PODC'05, July 17–20, 2005, Las Vegas, Nevada, USA.

1. INTRODUCTION

In a large scale peer-to-peer overlay network, physical entities that reside on different physical nodes communicate with each other using pseudonyms or logical identities. In the absence of direct physical knowledge of a remote entity or a certification by a central authority that a particular identity resides in a particular node, an entity can appear in the system under different names or *counterfeit* identities. Counterfeit identities are problematic in a peer-to-peer system because they can prevent entities from performing a remote operation, such as saving a file, multiple times to increase availability. An entity might select different identities to perform an operation, but these identities can all reside on the same corrupt entity resulting in a loss of redundancy. Counterfeit identities can also prevent the formation of reliable reputation-based recommendation systems. An entity that can create counterfeit identities can also create identities with fake reputations which makes reputations meaningless. Douceur [4] calls the forging of multiple identities a *Sybil attack.*

In this paper we study ways to restrict or prevent the damage that can result from corrupt entities performing *Sybil attacks.* In other words, we are interested in mechanisms to restrict the damage due to the creation of pseudonyms, while not relying on a centralized certifying authority or infrastructure with direct knowledge of entities in the system. While standard authentication techniques work well to prevent impersonation of existing identitities, they do not address the issues arising from the proliferation of pseudonyms. To our knowledge, the first work that studies counterfeit identities is the paper by Douceur [4]. He argues that under the strictest requirements, that is, in a fully distributed system without a central authority and in which entities communicate by broadcasting messages, the only means to limit the generation of multiple identities is by exploiting the fact that resources of individual entities are bounded.[1] Douceur argues that, by requiring any entity trying to establish an identity to dedicate a significant portion of its resources to this purpose, one could, at least theoretically, limit the number of identities that are forged by a corrupt entity. The three types of resources he considers are: computation, communication and storage.

[1]The absence of a central authority precludes the use of IP addresses to identify entities because they rely on the authority of ICANN. Furthermore, in real systems IP addresses can be spoofed and a host might be provided dynamic IP addresses by its ISP.

Our main observation is that the damage caused by Sybil attacks in Douceur's model is not only due to the fact that corrupt entities can forge multiple identities, but is also due to the fact that, in the absence of additional information, and for any two identities one of which is corrupt and has unbounded resources, one cannot conduct a test to determine that their entities are distinct. So, our goal need not necessarily be to test that any two particular identities are distinct, but rather to test that amongst a group of identities, a large enough subset of them resides on a set of distinct entities. Realizing such a test would allow the remote execution of remote operations and therefore circumvent the harm done by Sybil attacks. An example illustrates this point. Assume that one can divide identities into separate groups such that two identities in different groups are distinct, but two identities in the same group are not necessarily distinct. For concreteness, also assume that there is only one corrupt entity in the system. Under these assumptions, if an entity asks n entities in one group to perform an operation and another n entities in another distinct group to perform the same operation, it can be guaranteed that at least n distinct entities performed the operation even though it cannot tell which ones they are. If the operation consists of saving a file, the entity can be assured that there are enough correct replicas of the file in the system. The goal of this paper it to show that such a test is possible in real systems and to explore conditions under which such a test can be made accurately.

We develop our work by exploiting an ingredient that was eliminated by the strong assumptions of Douceur, namely that entities have certain physical properties, in particular, locations. Most of the real distributed systems we can imagine are in some way embedded in space with geometric properties. Moreover, the entities in the system have (at any moment in time) their own physical locations and no two entities share the exact same location at any moment. In general, we can assume that the underlying space (whose points include all the participants in the protocol) has a geometric structure of standard d-dimensional Euclidean space $\mathbb{R}^d$ or sphere $\mathbb{S}^d$. For sound or radio communication these assumptions are quite realistic (even though accuracy of measurement is always an issue) and they have already been exploited for secure location verification [6], while in the case of Internet-based overlay networks they are justified by recent work on estimating network distances [7] (we further discuss these assumptions as they relate to the Internet in Section 6). We distill our assumptions into the following:

1. the actual distances between pairs of entities at least approximately satisfy the metric properties (symmetry, definiteness, triangle inequality), and

2. the transfer of a message back and forth between two identities takes time that is lower-bounded by a (non-decreasing) function of the distance between the two entities on which they reside.

We do not assume that logical identities are always honest, and we place no bounds on the computational resources of corrupt entities. However, since each entity is located at a point in a geometric space, its communication with the rest of the identities in the protocol is restricted by the geometry of the space. In particular, a simple assumption of finite message propagation implies our second assumption above: the time in which a message is transmitted from point x to point y gives an upper bound on the distance $d(x, y)$ between the two points.

1.1 An example

To illustrate how physical locations can be used to provide a test of distinctness, consider two correct entities A and B at a distance d from each other. Assume that there is only one corrupt entity C in the system, that C has unbounded resources, and that C is within a radius of $d/2$ from A. Under these conditions, C can forge an unbounded number of identities, but all of these forged identities cannot pretend to be at a distance less than $d/2$ from B. In fact, for each identity c of C, one can request from A and B an upper bound on their distances to c. These distances can be obtained by having A and B broadcast probe messages to c and measure the time it takes to receive a reply from c. Since the distance from C to A is less than $d/2$ and the distance from A to B is d, it follows from the triangle inequality that the roundtrip time of probes sent from B to C (under any of its pseudonyms) will always indicate a distance that is larger than $d/2$ and therefore none of C's identities can prove that they are within radius $d/2$ from B. Using this test of distinctness, an entity can require that a remote operation be executed by n identities that can prove that they are within a radius of $d/2$ from A and another n identities that can prove that they are within a radius of $d/2$ from B and therefore be guaranteed that enough correct entities executed the operation. One can use the distance between A and B and their distance from c as a certificate of identity of c. Such a certificate allows one to determine that two identities are distinct if one certificate shows a distance smaller than $d/2$ to A and the other shows a distance smaller than $d/2$ from B. These certificates can be signed by A and B and later be presented to a different entity to prove distinctness. Note that such certificates give sufficient but not necessary conditions for distinguishing identities.

We should emphasize that A and B in the example above are not the same as a *centralized* certifying authority (we discuss this point further in Section 6). In fact, A and B's knowledge of C or its forged identities is obtained solely through *remote interaction* with C's various identities and with each other and the assumption that they are both honest (which we will not require in general), whereas a centralized certifying authority requires some form of direct knowledge of C. Also, note that A and B need not know each other's location, they only need to know the distance between them and that they are both honest.

1.2 Paper Outline

The goal of this paper is to study various scenarios under which entities such as A and B in the example above can be used to significantly restrict the types of counterfeit identities by a corrupt entity and therefore eliminate the harm caused by Sybil attacks. We show that one can construct certificates that are much more powerful than the one suggested above and that can be used under stronger adversarial conditions. The rest of the paper is organized as follows. Section 2 defines our system models, with additional discussion on the models available in Section 6. Section 3 summarizes our results and contributions, Section 4 discusses related work, and Section 5 present details of our results for the various models.

2. SYSTEM MODEL

We consider a system consisting of a set $\mathcal{B}$ of n *beacons* and a set of $\mathcal{A}$ of *applicants*. The set $\mathcal{A} \cup \mathcal{B}$ is the set of *participants*. We assume the participants are points in either the standard d-dimensional Euclidean space $\mathbb{R}^d$ or the d-dimensional unit sphere $\mathbb{S}^d$. In making statements that hold for both $\mathbb{R}^d$ and $\mathbb{S}^d$, we refer to the space as X. We denote by ρ the metric in X, that is, if $x, y \in X$, then $\rho(x, y)$ is the distance between x and y. We assume that participants are not mobile and that their locations are fixed.

2.1 Communication

Beacons communicate with each other and with applicants by exchanging messages. We distinguish two models for the messages transmitted by the participants: *broadcast* and *point-to-point*.

2.2 Failures

Some *applicants* and some beacons might be faulty (or corrupt). We assume that no more than f beacons are corrupt. The remaining beacons are correct (or honest). A corrupt beacon can report fake distances to any participant, and these distances can be smaller or larger than the actual distance to the participant. An applicant can also be corrupt and it can delay its responses for probes from the beacons therefore making it appear farther away than it really is. We consider cases in which applicants might collude and cases in which corrupt applicants do not collude. To strengthen our adversarial model, we assume that corrupt applicants and beacons know the locations of all beacons and that correct beacons do not know each other's locations other than what can be implied from the time it takes to receive replies from probes. We assume that the distance between two correct entities is a non-decreasing function of the roundtrip delay between them (we discuss this and our other assumptions in detail in Section 6). We use μ for the distances as *measured* by exchanging messages between the points. Thus $\mu(A, B)$ is the distance A can deduce from the roundtrip time of a message transmitted from A to B and back. For correct participants A and B, we assume that $\mu(A, B) = \rho(A, B)$ that is, the distance can be accurately measured by observing the roundtrip delay. Note that in the presence of faulty participants μ is not necessarily symmetric. For a participant A in the system, we denote by $x(A)$ the location of A in the underlying geometric space.

2.3 Geometric Certificates

An applicant can request a *geometric certificate* from a set of beacons. When an applicant requests a geometric certificate, the beacons and the applicant execute a protocol that might require the applicant (as well as other beacons) to respond to probe messages. The protocol might also require the applicant to send probe messages to the beacons and report distances to various beacons. The result of these exchanges will be a geometric certificate: a set of distance values between the beacons and the applicants that are signed by the beacons as well as the applicant.

2.4 Distinctness Test

A distinctness test is a function $D : \mathcal{C} \times \mathcal{C} \mapsto \{\mathbf{true}, \mathbf{unknown}\}$ that assigns to a pair of geometric certificates a value in the set $\{\mathbf{true}, \mathbf{unknown}\}$. If $D(c_1, c_2) = \mathbf{true}$, then the entities that obtained these certificates are distinct.

2.5 Synchrony and Reliability

We assume that the system is asynchronous and that message transmission is not reliable, but that there are periods of time during which message transmission is synchronous and reliable. We assume that for large enough time intervals, the system will enter a synchronous period. While these assumptions do not really simplify online communication between peers in an overlay network, because peers cannot wait for the periods of synchrony, they are necessary for establishing geometric certificates. The idea is to have participants probe each other for a somewhat long period of time in order to get an accurate measure of distance. In fact, we expect that the measurements during periods of synchrony (low congestion periods) accurately reflect the distances between correct participants. Once certificates are obtained, we do not require any synchrony assumptions for communications between applicants. Beacons have local clocks and the rate of drift of these clocks is small enough so that clock drift is negligible during the time it takes to establish a certificate.

3. CONTRIBUTIONS AND SUMMARY OF RESULTS

The main contribution of this work is to show that it is possible to *remotely* issue certificates that can be used to test the distinctness of identities. To our knowledge this is the first work that shows that remote anonymous certification of identity is possible under adversarial conditions.

The following is the summary of our results. We present geometric certification protocols, which issue compact and easily-checkable certificates to applicants. Given two certified entities, a distinctness test may be performed, and if the two entities' geometric locations are distinguishable from the point of view of the beacons that participated in the certification protocol, the distinctness test will succeed and certify that the two entities are indeed distinct. The certification protocols we present work for several different settings, including the following (in all cases, we assume the number of beacons is at least $d + 1$, where d is the dimension of the space; also, unless otherwise specified, the applicant entity or entities should be in the convex hull of the certifying beacon set; finally, beacons' messages are always broadcast):

(1) honest participants: no restrictions;
(2) corrupt applicant: either applicant in the convex hull of beacons (in $\mathbb{R}^d$), or a sufficient set of beacons (in $\mathbb{S}^d$) without restriction on the applicant's location (for an exact definition, see Section 5.1.2);
(3) multiple colluding entities: broadcast message model;
(4) multiple colluding entities (at most d of them): point-to-point (arbitrary) applicant message model;
(5) up to f corrupt beacons, at least $f + d + 1$ correct: single corrupt applicant entity, or multiple colluding applicant entities.

4. RELATED WORK

Ng and Zhang [7] model the Internet as a geometric space by using measurements of round-trip delay for ICMP ping messages between a set of known hosts *probes* and several sets of *targets*. They assign the targets to points in a coordinate system, by defining each coordinate of a node as its distance from one of the probes. This *embedding* into a low-dimensional Euclidean space allows them to derive simple lower and upper bounds on distances between targets from their probe-target measurements by using the triangle inequality.

Kleinberg et al. [5] design algorithms that try to infer a complete distance matrix of a finite set of points, given only the distances from a small number of selected points (*beacons* in their terminology) to every other point. They show that most of the distances can actually be reconstructed even from such limited data, but also that arbitrary distortion of a certain fraction of all distances is unavoidable. They use some of the powerful recent results on metric embeddings and provide very general algorithms, however their results do not seem to have immediate applications to the Sybil attack problem.

Newsome et al. [6] study the Sybil attack in the context of sensor networks, and so their approach relies heavily on strong restrictions of computational, communication and memory resources available to the nodes. While these restrictions are realistic for certain sensor networks, their techniques do not exploit locational properties of the nodes.

The closest to our approach in the existing literature seems to be the paper of Sastry et al. [8]. They do study protocols for establishing identity of nodes based on their location. Their methods only use single beacons (*verifiers* in their terminology). This doesn't allow them to determine exact locations of nodes and reduces most of the problems to simple covering problems. Also, they do not consider adversarial conditions such as faulty beacons or collusions by applicants.

5. GEOMETRIC CERTIFICATION PROTOCOLS

In this section we present our results under various system assumptions. For each set of assumptions, we state our results in the form of a theorem that specifies conditions under which a participant (or group of participants) is incapable of pretending to be in a location other than the real location of the participant or one of the group members. We say that a participant (or a group of participants) simulates a point, if it can make all its communications appear to come from the point.

These results can be readily used to construct geometric certificates for the applicant. In each case, a certificate would consist of the set of measurements that is sufficient to uniquely identify the location of an entity, and a test of distinctness is simply a comparison between the two locations defined by two certificates.

All our results are stated assuming the distance between correct participants is accurately measured using roundtrip delays (as explained in Section 2). These results can be extended to the case in which measurements are not accurate. For that case, the statements of the theorems will change to specify conditions under which an applicant is incapable of pretending to be outside of a well-defined neighborhood of its actual location. In the presence of inaccuracies, a certificate consists of the measurements that establish a neighborhood of the applicant's location, and a test of distinctness is simply the test of disjointness of two such neighborhoods. While we do not describe such protocols here, our results can be generalized to account for small inaccuracies (as outlined in Section 5.4).

In our model, we assume only that the distances between beacons can be calculated, while the locations of beacons are unknown. Given a distance matrix M_d whose entries are the pairwise distances between points in a geometric space, it is possible to find a set of points expressed in an orthonormal coordinate system and whose distance matrix is identical to M_d [2, 3]. If all beacons are correct, these methods can be used to transform a distance matrix representation into a coordinate system representation. In the presence of faulty beacons, the calculated distance matrix might not be realizable in a geometric space and a coordinate representation consistent with all the beacons might not be possible. Still, in the presence of faulty beacons, the distance matrix is realizable if it is restricted to the set of correct beacons. So, our goal would be to find a realization that is guaranteed to be consistent with the set of correct beacons. Assuming that the set of correct beacons is in general position (that is, no $(d+1)$-subset is contained in a d-dimensional hyperplane, and no $(d+2)$-subset is contained in a d-sphere), this can be easily achieved by considering either all sets of $d+1$ or all sets of $d+f+1$ beacons (depending on which of the two families is smaller). In the first case, we use each $(d+1)$-set to build a coordinate representation and then check if there are another f beacons consistent with this representation. In the second case, we look for a consistent $(d+f+1)$-set of beacons. In case such a set is found, it must contain at least $d+1$ correct beacons, therefore the coordinate representation defined by this set is consistent with all the correct beacons and every beacon inconsistent with this representation can be discarded as faulty.

It is important to note that, while the procedure described above is expensive—being exponential in the (usually small constant) d, and including a verification of the positive-semidefinitness of a matrix—it is only performed once for an applicant to establish the certificate. The size of the certificate itself is small, and the test of distinctness efficient.

5.1 Honest beacons with known locations

5.1.1 Trilateration in an honest world

If all participants in the protocol are honest, then the problem is easy. To determine the exact location of a point $x(A)$ in d-dimensional space, it is enough to know all the distances $\rho(x(A), x(B_i))$ between $x(A)$ and $d+1$ other affinely independent points $x(B_1), \dots x(B_{d+1})$. With this information, the point $x(A)$ can be reconstructed as follows: let S_i be the sphere of diameter $\rho(x(A), x(B_i))$ around $x(B_i)$. The point $x(A)$ belongs to S_i for every i. A sphere with center $c = (c_1, \dots, c_d)$ and radius r is the set of all points $x = (x_1, \dots, x_d)$ that satisfy the equation $\sum_i (x_i - c_i)^2 - r^2 = 0$. Equating the left-hand sides of the equations for S_i and S_j gives a linear equation in x_i, thus the intersection of two spheres belongs to a hyperplane. Since we assume general position, each pair S_1, S_i defines a hyperplane, which we denote by H_i. Since $S_1 \cap S_i \subseteq H_i$, it follows that $x(A) \in \cap_i H_i$, and we can determine $x(A)$ by solving a linear system.

5.1.2 Trilateration against cheaters

In the situation where the applicant may cheat by pretending to be at a different location, the protocol should compute the applicant's position or detect the cheating. We first discuss the case where a single point attempts to cheat without colluding with other entities.

Consider an applicant at A that attempts to impersonate a point $x' \neq x(A)$. The applicant contacts $d+1$ beacons $B_1, \ldots, B_{d+1}$ and exchanges a message with each of them. Let $\mu_i = \mu(B_i, A)$. If A can successfully impersonate x', then $\mu(B_i, A) = \rho(x(B_i), x')$ for every i. Since $\mu(B_i, A) \geq \rho(x(B_i), x(A))$, it follows that $\rho(x(B_i), x(A)) \leq \rho(x(B_i), x')$ for every i, that is, $x' \in D_1 \cap \cdots \cap D_k$, where $D_i = \mathbb{B}(x(B_i), \mu(B_i, A))$, the ball of radius $\mu(B_i, A)$ around B_i. For a set Z, we use $\operatorname{int} Z$ to denote its interior, and $\operatorname{conv} Z$ its convex hull.

THEOREM 1. *Let $X = \mathbb{R}^d$. Let B_i be a beacon with $x(B_i) = b_i$ for each $i = 1, \ldots, d+1$. If $\{b_1, \ldots, b_{d+1}\}$ is affinely independent and $x' \in \operatorname{int}\operatorname{conv}\{b_1, \ldots, b_{d+1}\}$, then x' cannot be simulated by any other point.*

PROOF. First, the set S of all points that can simulate x' can be written as $S = \{x \mid \forall i\ \rho(b_i, x) \leq \rho(b_i, x')\}$. If $S \neq \emptyset$, take $x^* \in S$. Since S is an intersection of closed balls, it is convex and so $x_\lambda = \lambda x^* + (1-\lambda)x' \in S$ for all $0 \leq \lambda \leq 1$. Take $\lambda > 0$ small enough that $x_\lambda \in \operatorname{conv}\{b_1, \ldots, b_{d+1}\}$. If we can show that for some i^*, $\rho(b_{i^*}, x_\lambda) > \rho(b_{i^*}, x')$, it will follow that $x_\lambda \notin S$, and the proof by contradiction will be complete. So assume that $\rho(b_i, x_\lambda) \leq \rho(b_i, x')$ for all i. Let H be the hyperplane through $\frac{1}{2}(x' + x_\lambda)$ with normal vector $x_\lambda - x'$. H is exactly the set of points at equal distance to x' and x_λ. This implies that for every i, b_i is on the same side of H as x_λ, in other words, the hyperplane H separates x' from b_i for every i. This contradicts the fact that $x' \in \operatorname{conv}\{b_1, \ldots, b_{d+1}\}$. $\square$

If the underlying geometric space X is $\mathbb{R}^d$, then the theorem above gives necessary and sufficient conditions for a set of beacons to be able to detect a cheating point. If x' is not in the convex hull of the beacon set, it may be impossible to detect cheating.

However, on the sphere $\mathbb{S}^d$, the situation is different, and in fact much better. Note first that on a sphere, the distance between a pair of points is measured along a geodesic curve (great circle) that connects the pair. The notion of convexity can then also be defined for the sphere. Given two points $x, y \in \mathbb{S}^d$, we would like to define their convex hull as the set of all points on the shorter segment of the geodesic between x and y. This works because there is a unique x-y geodesic as long as x and y are not antipodal points. (If x and y are antipodal, then there are infinitely many geodesics between x and y, and in fact every point on the sphere lies on one of them.) Now consider a finite set of points $X = \{x_1, \ldots, x_k\}$ on the d-dimensional sphere, such that X is contained in some open half-sphere (this ensures that no two points of X are antipodal, and more generally that our definition is useful). We define the convex hull $\operatorname{conv}(X)$ inductively: $\operatorname{conv}(\{x_1, x_2\})$ is the (shorter) segment of the geodesic between x_1 and x_2. Then suppose we have defined $C_i = \operatorname{conv}(\{x_1, \ldots, x_i\})$. We define $C_{i+1} = \operatorname{conv}(\{x_1, \ldots, x_i, x_{i+1}\})$ to be the union of segments between x_{i+1} and y, where y ranges over all the points in C_i. (The same definition can be stated for the Euclidean space and results in the standard notion of convexity there.)

As an example of our claim that the situation on the sphere is even better than in the Euclidean space, consider the following theorem.

THEOREM 2. *Let $X = \mathbb{S}^{d-1}$. Let $b_1, \ldots, b_{2d}$ (the locations of beacons $B_1, \ldots B_{2d}$) be the points defined by the coordinate unit vectors $+e_1, -e_1, +e_2, -e_2, \ldots, +e_d, -e_d$ in both orientations (that is, if $\mathbb{S}^{d-1}$ is considered as a subset of $\mathbb{R}^d$, the beacons are at the vertices of the polar of the inscribed d-dimensional cube). Then no point $x' \in \mathbb{S}^{d-1}$ can be simulated by any other point.*

PROOF. (Sketch.) First, assume without loss of generality that x' is in the interior of the positive orthant. (By symmetry x' can be assumed to be in the positive orthant, and if it is on the boundary, then we can restrict the discussion to $\mathbb{S}^{d-2}$ and continue the proof there using only the appropriate subset of beacons.) We can now use only the d beacons located at $+e_1, \ldots, +e_d$, We claim that for any $x^* \neq x'$, there exists an i such that $\rho(b_i, x^*) > \rho(b_i, x')$. For the most interesting case, where the simulating point is also located in the positive orthant, the proof is the same as that for Euclidean space in Theorem 1 with the definitions of convexity, distance and hyperplane modified to fit the spherical geometry of $\mathbb{S}^{d-1}$. $\square$

The boundedness of the sphere makes it possible to use a single "universal" set of beacons to distinguish any point from any other. We do not claim that we can always place beacons exactly at the locations used in Theorem 2. However, this theorem gives a sufficient condition to ensure that every point in the space is contained in the convex hull of some set of beacons, and therefore cannot be simulated by any other point. (As long as the number of beacons is finite, this cannot be done in Euclidean space because the applicant can always be far enough to be outside of the convex hull of the beacons.) In practice, we may use any appropriate beacon set, but distinguishability may be more difficult to guarantee if the beacons used are not all within a single half-sphere because then we must be very careful about convexity.

5.2 Multiple colluding entities

We have seen that it is impossible for a single applicant at point x^* to impersonate any other point x' in the convex hull of a sufficiently large set of active beacons. However, the proof was based on the fact that the distance from x^* to some beacon would have to be greater than from x' to the same beacon and so x^* couldn't return messages in time. If several entities located at different points collude to jointly impersonate another point, our protocols from Section 5.1.2 don't work anymore. In fact, in this setting there is a significant difference between the problem under the broadcast and point-to-point models.

5.2.1 Broadcast messages

In the broadcast model the applicant cannot send a message to a single recipient. Instead, every message sent is broadcast and thus received by every other entity (or at least every entity expecting a message). More precisely, every message sent by an applicant A at time t is received by every beacon B_i at time $t + \rho(x(A), x(B_i))$.

THEOREM 3. *Let x' be a point surrounded by an independent set of beacons $\{B_1, \ldots, B_{d+1}\}$, either in the sense of Theorem 1 in $\mathbb{R}^d$ or in the sense of Theorem 2 in $\mathbb{S}^d$. In the broadcast model, x' cannot be simulated by any set $\{A_1, \ldots, A_k\}$ of entities unless $x(A_i) = x'$ for some i.*

In the proof of the theorem, we simply use A to denote the entity of the applicant. The first reason for this is that no beacon can tell which A_i sends the message just by the content of the message since $A_1, \ldots, A_k$ are in collusion. The second reason is that the broadcast model ensures that each message sent by any of the entities $A_1, \ldots, A_k$ is received by all beacons.

Proof. The protocol begins by the beacon B_1 sending a probe at time 0. The applicant responds (broadcasting to all the beacons). Upon receiving the response, each beacon immediately forwards it to B_1. Now B_1 records $\mu(B_1, A)$ and the times at which it receives the responses from the other beacons. Given the distances $\rho(x(B_1), x(B_i))$, the beacon can deduce the time it took the message from A to arrive to each beacon by subracting $\mu(B_1, A)/2 + \rho(x(B_1), x(B_i))$ from the time t_i at which the response from B_i arrived. Therefore, the beacon B_1 can calculate the distance $\mu(A, B_i)$ for each i. Now by Theorem 1 for $\mathbb{R}^d$ or by Theorem 2 for $\mathbb{S}^d$, the point x' cannot be simulated by any other point. □

5.2.2 Point-to-point messages

The case in which an applicant can send a message to any single beacon appears to be substantially more difficult in the case of colluding entities. For example, our protocols from the previous section fail because the simulating entities $A_1, \ldots, A_k$ can reply to the beacons selectively, so that A_i sends a message to B_j only if A_i is closer to B_j than x'. Thus a point x' can be simulated whenever for each beacon B_i there exists an A_j such that $\rho(x(A_j), x(B_i)) \leq \rho(x', x(B_i))$.

Theorem 4. *Let $X = \mathbb{R}^d$ and let x' be a point in the convex hull of $d+1$ affinely independent beacons. Then no d entities can simulate x' unless one of them is located at x'.*

Proof. (Sketch.) We describe the protocol and give the proof for $d = 2$, because it is easy to visualize and contains all the necessary ideas.

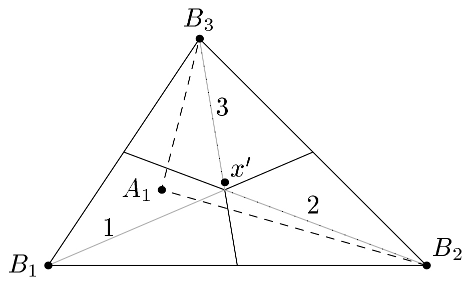
The protocol consists of two rounds. In the first round, the beacons compute x' from the values of $\mu(B_i, A)$, $i = 1, \ldots, d+1$. In the second round, the beacons synchronize clocks and broadcast messages $M_1, \ldots, M_{d+1}$. The message M_i is sent by beacon B_i, so that at time t_0 all $d+1$ messages simultaneously reach x'. The messages should be impossible to forge. A simple solution is that each beacon sends a random message to the applicant. The applicant must then immediately forward all three messages to each beacon. (The beacons later verify that the forwarded messages are indeed the messages they sent to the applicant.)

Consider the situation for $d = 2$, where there are three beacons (B_1, B_2 and B_3) and two entities, (A_1 and A_2). In order that A_1 and A_2 successfully simulate x', each beacon B_j must receive a message forwarded from either A_1 or A_2 on time (no later than by time $t_0 + \rho(x', x(B_j))$). From the definition of the protocol, if A_i can forward the required message to B_j on time, it must be true that

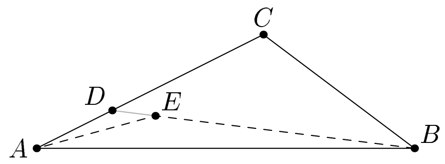
$$\rho(x(B_k), x') + \rho(x', x(B_j)) \leq \rho(x(B_k), x(A_i)) + \rho(x(A_i), x(B_j))$$

for all $k \neq j$. A necessary condition for this is (see Figure 1(a)) that A_i lies on the same side as B_j of the two lines defined by (x', B_k) for $k \neq j$. (For a proof of this, consider Figure 1(b)).

This proof generalizes quite directly to the case of arbitrary dimension d. Let $S = \mathrm{conv}(\{x(B_1), \ldots, x(B_{d+1})\})$ be the convex hull of an affinely independent set of $d+1$ beacon points. Notice that S is a d-simplex. For each pair of beacons B_i, B_j, let $H_{i,j}$ be the hyperplane spanned by x'



(a) A_1 cannot cheat B_3 because A_1 doesn't lie in region 3



(b) $|AE| + |EB| \leq |AC| + |CB|$

Figure 1: Proof that two entities cannot cheat.

and all the beacons except for B_i and B_j. The $\binom{d+1}{2}$ hyperplanes $H_{i,j}$ all meet at x', and the intersection of $H_{i,j}$ with S can be written as a union of d subsets, all defined by the intersections of $H_{i,j}$ with the other hyperplanes. (For an illustration in the case $d = 3$, consider Figure 5.2.2.) Of the d parts into which $H_{i,j}$ is broken up by the boundary of S and the other hyperplanes, $d-1$ contain a beacon point (recall that $H_{i,j}$ passes through $d-1$ of the beacon points). The d-th one does not and this is the one we focus on. Denote by $F_{i,j}$ this subset of $H_{i,j}$ bounded by the intersections of $H_{i,j}$ with the other hyperplanes and by the boundary of S. Then the $\binom{d+1}{2}$ hyperplane segments $F_{i,j}$ break up S into $d+1$ pieces, each containing exactly one of the beacon points. Call these sets T_i.

Since there are only d colluding points and $d+1$ beacons,

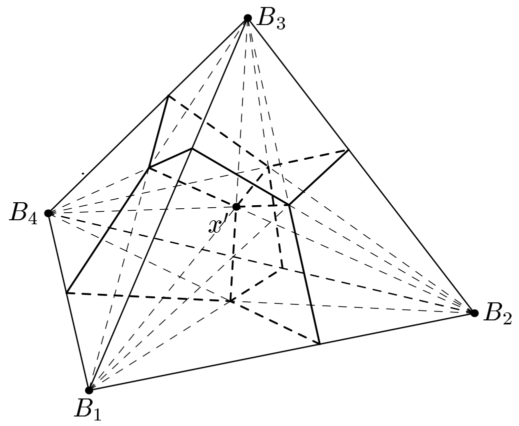


Figure 2: In three dimensions, three entities cannot cheat.

there exists an i such that the interior of S_i contains no colluding points. Now we can proceed to the second part of the proof and consider the simplex S_i' spanned by x' and all the beacons except for B_i:

$$S_i' = \mathrm{conv}(\{x', x(B_1), \ldots, x(B_{d+1})\} \setminus \{x(B_i)\}).$$

For any any colluding entity A_ℓ, there exist two vertices of S_i', say, $x(B_j)$ and $x(B_k)$, such that A_ℓ is neither in $H_{i,j}$ nor in $H_{i,k}$. This then implies $\rho(x(B_j), x') + \rho(x', x(B_k)) < \rho(x(B_j), x(A_\ell)) + \rho(x(A_\ell), x(B_k))$, that is, A_ℓ cannot simulate the message sent from B_i to B_j. □

5.3 Trilateration with corrupt beacons

In the presence of faults, we cannot rely on the operation of any single beacon to work as specified by the protocols. We now show how to tolerate beacon failures.

THEOREM 5. *Let x' be the location claimed by the applicant A. Consider a set of $d+1+2f$ affinely independent beacons surrounding x', at most f of which are faulty. Then no applicant A can simulate x' unless $x(A) = x'$. A certificate for the applicant can be constructed in time* $\min\{\binom{d+1+2f}{d+1+f}, \binom{d+1+2f}{d+1}\}$.

A set of $d+1+2f$ beacons contains at most f faulty, and therefore at least $d+1+f$ correct beacons. The f faulty beacons may, together with at most d correct beacons determine an incorrect value for $x(A)$. However, there are more correct beacons and so the maximum subset of beacons agreeing on a location for x' is correct. Unfortunately, finding a maximum consistent subset of a set of n linear equalities is not only NP-hard, but also hard to approximate within an n^ϵ factor for some $\epsilon > 0$ [1].

PROOF. We check either all sets of $d+1$ or all sets of $d+f+1$ beacons (depending on which of the two families is smaller). In the first case, for every $(d+1)$-set of beacons, we find a candidate point for x' (by solving a linear equality system as in Section 5.1.1). Then for each beacon B, we check if $\mu(B, A) = \rho(x(B), x')$. If there are at least $d+1+f$ such beacons, then $x(A) = x'$.

In the second case, if each beacon in a set of $d+f+1$ beacons is consistent with x', then this set of beacons defines a unique point, which must be the location of the applicant, because a $(d+1)$-subset of this set consists of correct beacons, and so also defines the actual location of A. □

THEOREM 6. *Let x' be the location claimed by the applicant A. Consider a set of $d+1+2f$ beacons surrounding x', at most f of which are faulty. Then no set of entities $\{A_1, \ldots, A_k\}$ can simulate x' in the broadcast model unless $x(A_i) = x'$ for some i.*

PROOF. As in the proof of Theorem 5, if A is honest, the set S will have a unique element, and more beacons will agree on this point than on any other. The required distances are computed as in the proof of Theorem 3. □

5.4 Inaccuracies

In this section we very briefly describe a generalization of the problem that allows inaccuracies in measured distances. We only consider the simplest case, where measured distance may vary by a small constant fraction (known in advance) from the actual distance. This assumption may not be as

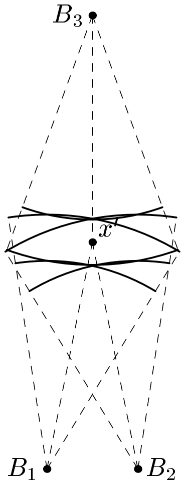


Figure 3: The small angle case.

restrictive as it appears, especially in view of our suggestion in Section 2.5 that each distance be measured more than once over a period of time to ensure accuracy (for example, if the measurements deviate from the true value according to a reasonable probability distribution, the smallest of the multiple measurements will generally tend to the true value fairly quickly).

5.4.1 Inaccurate distance measurements

To account for the inaccuracies in measuring the distance, we consider the variant of the problem where the measured distance $\mu(B_i, A)$ is allowed to exceed the actual reported distance by a small multiplicative factor: as long as there exists a point x' that satisfies $\rho(x(B_i), x') \le \mu(B_i, A) \le (1+\epsilon)\rho(x(B_i), x')$ for each i, the protocol should not reject the applicant.

The problem of validating an applicant becomes equivalent to identifying an intersection of thin (because we assume ϵ to be small relative to the measured distances) spherical shells (one around each of the beacons).

The goal of all our protocols is to distinguish between different applicants. Therefore a natural measure of how badly a protocol fails might be the smallest distance between points that cannot be reliably distinguished using the protocol. Now imagine a region E such that no two points in E can be reliably distinguished. Since E is an intersection of shells around beacons and the thickness of each shell is small compared to its radius, we may think of E as bounded by almost straight planar surfaces. Consider as an example the two dimensional case, and focus first on just two of the beacons, B_1 and B_2. The straight line segments from their locations to the applicant's location meet at an angle, say θ. If θ is close to the right angle, then the indistinguishability region is close to a square (actually, two disjoint squares, because two circles around B_1 and B_2 intersect in two points). If θ is very small, the indistinguishability region looks more like a thin parallelogram, and in such a case it can happen that two points a large distance apart cannot be distinguished. When the third beacon is included, it may still be the case that the indistinguishability region has a large diameter.

For this case, if the shell boundaries are replaced by straight lines, we see that the distance between the two points farthest apart in the indistinguishability region is at most $D = 2d\epsilon\frac{\cos(\theta/2)}{\sin\theta}$, where d is the distance between B_1 and x'. In other words, the diameter of the indistinguishability region may increase proportionally to $1/\sin\theta$. We defer a more detailed analysis of this problem to the full version of this paper.

6. DISCUSSION OF MODEL

6.1 Certificates

It is important to realize that the set of beacons in our model is not the same as a central certifying authority. In fact, all we need to assume about the beacons is that they are distinct and that a certain number of them are correct. In principle, a given entity can establish the distinctness of an initial set of beacons by using some of the resource-consuming challenge-response described in [4] without requiring any certifying authority. The assumption that a certain proportion of beacons chosen at random is correct is a system assumption for we cannot expect a system with an arbitrary number of faulty entities to be able to function.

Once an initial set of beacons is established, our results show that they can be used *remotely* to establish the distinctness of identities created by entities with unbounded computing power. This shows that the following lemma from [4] (his notation is different from ours, but should be clear from the context) does not hold once the geometric properties of communication are considered.

LEMMA 7. *Lemma 4 [4] If the correct entities in set C do not coordinate time intervals during which they accept identities, and if local entity l accepts any identity vouched for by q accepted identities, then even a minimally capable faulty entity f can present $g = \lfloor |C|/q \rfloor$ distinct identities to l.*

This lemma basically says that accepted identities cannot be used to accept further entities. We showed that, if we take the geometric properties of communication into account, we can use accepted identities to accept additional entities. In fact, in one of our results we showed that a set of $d + 1 + 2f$, at most f of which are faulty, can prevent one faulty entity e_{faulty} in their convex hull from presenting distinct identities even if e_{faulty} has unbounded resources. This result is achieved without assuming a central authority.

In practice, the beacons can be certified by a central certifying authority to bootstrap the system. Once a set of beacons is certified, it can be used to provide certificates remotely. In that case, an applicant that wants to obtain a certificate from the set of beacons would identify beacons that have valid public certificates obtained from the central authority. Then, the applicant can initiate a geometric certificate request which will result in the beacons probing the applicant as explained in the various protocols we presented. These probes will be started by multiple beacons to obtain the distances as required by the protocols. At the end of the probing period, the beacons will present the applicant with pieces of the geometric certificate (distances from beacon to applicant or location of applicant as calculated by a beacon) that the applicant can put together to obtain the geometric certificate.

6.2 Limitation of distinctness tests

Our approach is conservative. The distinctness tests we propose are sufficient but not necessary to establish distinctness of two identities. For example, different machines that reside on the same LAN would appear to be at the same location to remote beacons (assuming these machines do not route through each other). This does not mean that we can only use one machine from a set of machines that appear to be at the same location. As our introductory example shows, to execute a remote operation, one can choose multiple groups of machines such that machines in each group appear to be at the same location. If the number of groups exceed the assumed upper bound on faulty entities, then the collection of groups would be guaranteed to contain at least as many entities as in the smallest amongst them.

The distinctness tests we presented assume that the entity under consideration is in the convex hull of the beacons in the system. If an entity is outside the convex hull of the beacons, then most of the theorems we prove do not hold. It is reasonable to question whether the convex hull condition is only of theoretical interest and if anything can be done if an entity is not in the convex hull of the beacons. The answer to the first part of the question would depend on the actual network under consideration. We can expect, and that needs to be verified by further study, that if the network is dense and beacons are evenly distributed in the network, then it should not be hard to find a set of beacons that surround and entity. Even if the points are not in the convex hull of the beacons, it seems possible that one can generalize the introductory example

6.3 Communication

A number of our results assume that the only mode of communication available to applicants is the broadcast of a message to all other participants. These results have limited applicability because in actual overlay networks peers communicate by a combination of point-to-point communication and broadcast. It is interesting to note, though, that restricting communication to broadcast makes it harder for faulty applicants to benefit from colluding, whereas in Douceur's argument broadcast makes is harder to prevent Sybil attacks. The reason is that we assume that our broadcast primitives is an indivisible operation that cannot be split into multiple point-to-point operations. If the broadcast operation can be implemented using multiple point-to-point operations, then our result for tolerating colluding entities in the broadcast mode would no longer hold.

6.4 Accuracy of measured distances

Ng and Zhang [7] show that on the Internet the roundtrip delays can be used to measure distances between entities if enough measurements are taken and the minimum amongst the measured delay is used as the distance measure. These measurements were done using ICMP ping messages. In our model, communication is done in an overlay network that does not necessarily exhibit the same delay characteristics as those of the Internet. Nonetheless, we can expect that in periods of low congestion, the distances will reflect the underlying network distances. In our work, establishing a geometric certificate can be done over a period of time and multiple measurements can be taken and the smallest times be included in the certificates. If the participants belong to common congestion zones (which can be corollated with time zones), then we can expect that the minimal delay measured by participants will exhibit metric characteristics. Nevertheless, studying delay characteristics in Internet-based overlay networks is a subject that needs further study and our work is based on the assumption that these characteristics are similar to those of the Internet.

6.5 Adversary model

Throughout, we assumed that the corrupt entities have unbounded computation power. Our protocols implicitly assume that an entity cannot anticipate probe messages and send replies before the receipt of the actual probes. This assumption can be easily enforced by using the standard technique in which each probe message includes a randomly generated string that only the sender knows and that must also be included in the reply. This way, an entity would have only a very small probability of successfully being able to reply before receiving a probe.

The adversary model is particularly important for the accuracy of measured distances. In fact, a corrupt entity that is used to route messages between non-corrupt beacons can artificially increase the distance between them as well as between corrupt entities and beacons. Later, the calculated distance could be shortened which violates a main assumption of our model. We do not have a fully satisfactory solution to this problem, but we have two approaches to deal with it. The first approach applies to peer-to-peer systems that exhibit locality characteristics. In such systems, the distance between nodes is proportional to the actual network distance between the nodes. If the overlay network exhibits locality characteristics, we can calculate the network-distances between beacons directly without going through the overlay network and therefore without risking that the routing is compromised (assuming the routing on the underlying network cannot be easily compromised). These distances will be smaller than the distances on the overlay network, but one could then use solutions that tolerate inaccuracies in the measured distances. The second approach makes limiting assumptions on the disruption power of the adversary. If we assume that any two nodes are connected by a path that does not go through a corrupt node, then we can use multiple paths to calculate the distance between two nodes; The shortest amongst the calculated distances would be chosen as the distance between two nodes.

Another potential difficulty can be cause by corrupt nodes trying to flood the network with message in order to prevent accurate measurements of distances. Dealing with such denial of service attacks is beyond the scope of this paper.

7. CONCLUSION

We have shown that it is possible to exploit the geometric properties of message transmission delay in order to reduce the effects of Sybil attacks. We believe that a lot more work is still needed to make this work of more practical value. In particular, we believe that a generalization of the introductory example that directly exploits the triangular inequality has a good chance of leading to solutions that can be used in practice.

8. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their comments.

9. REFERENCES

[1] E. Amaldi and V. Kann. The complexity and approximability of finding maximum feasible subsets of linear relations. *Theoretical Computer Science*, 147:181–210, 1995.

[2] L. Blumenthal. *Theory and applications of distance geometry.* Clarendon Press, 1953.

[3] M. Deza and M. Laurent. *Geometry of cuts and metrics.* Springer, 1997.

[4] J. Douceur. The Sybil attack. In *Proc. of IPTPS*, pages 251–260, 2002.

[5] J. Kleinberg, A. Slivkins, and T. Wexler. Triangulation and embedding using small sets of beacons. In *Proc. of the IEEE FOCS*, pages 444–453, 2004.

[6] J. Newsome, E. Shi, D. Song, and A. Perrig. The Sybil attack in sensor networks: analysis and defenses. In *Proc. of IPSN*, 2004.

[7] T. Ng and H. Zhang. Predicting Internet network distance with coordinates-based approaches. In *Proc. of INFOCOM*, 2002.

[8] N. Sastry, U. Shankar, and D. Wagner. Secure verification of location claims. In *Proc. of ACM WiSe*, 2003.