Street-Level Trust Semantics for Attribute Authentication

Tiffany Hyun-Jin Kim, Virgil Gligor, and Adrian Perrig

Carnegie Mellon University
Pittsburgh, PA. 15213
{hyunjin1,virgil,adrian}@ece.cmu.edu

Abstract. The problem of determining whether a receiver may safely *accept attributes* (e.g., identity, credentials, location) of *unknown senders* in various online social protocols is a special instance of a more general problem of establishing trust in interactive protocols. We introduce the notion of *interactive trust protocols* to illustrate the usefulness of *social collateral* in reducing the inherent trust asymmetry in large classes of online user interactions. We define a social collateral model that allows receivers to accept attributes from unknown senders based on explicit recommendations received from social relations. We use social collateral as a measure of both social relations and "tie strength" among individuals to provide different degrees of accountability when accepting attribute information from unknown senders. Our model is robust in the face of several specific attacks, such as impersonation and tie-strength-amplification attacks. Preliminary experiments with visualization of measured tie strength among users of a social network indicate that the model is usable by ordinary protocol participants.

1 Introduction

In many real-world social interactions, the authentication of someone's attributes is a crucial requirement. For example, accepting an invitation to a social event from an unknown person often requires an introduction that establishes that person's identity, and possibly credentials. If verified, that person's social connections may be sufficient to establish an identity that is suitable for the invitation protocol. In short, many social interactions in the physical world rely on one's ability to authenticate others' attributes. In these interactions, a receiver's friends, family members, or professional colleagues are often able to authenticate the attributes of a sender with whom they are acquainted but who may be unknown to the receiver. Typically a receiver accepts the attribute authentication of an unknown sender from his/her social relations conditionally, depending upon affirmative answers to the following two questions. First, is the receiver's social relation sufficiently strong (e.g., close friend, immediate-family member, colleague of long standing) to warrant the receiver's trust for the particular protocol? And second, does the receiver's social relation know the sender well enough to competently vouch for the authenticity of a particular attribute?

As social interactions migrate from the physical to the online world, it would be desirable that the authentication of an unknown sender's attributes would

proceed along the same lines as in a physical-world protocol and match a receiver's natural expectations. Users are less likely to make costly mistakes and accept unauthentic inputs, disclose private information, and fall victims of online scams if protection measures with which they are familiar in the physical world are supported in the online world. However, current online social networks do not use any form of social authentication from the physical world, and as a consequence they cannot guarantee the correspondence between an online and a physical-world identity. A typical example is a Facebook invitation, which cannot guarantee the physical identity of the issuer or even that the issuer exists in the physical world. Authenticating an individual's identity by examining a list of mutual Facebook "friends" provides inadequate identity authentication in practice [1, 2, 13], even for security-conscious individuals [23]. Furthermore, associations between an online identity and a public key, which is typically provided by identity certificates, are becoming more and more uncertain [3] – not just cumbersome to obtain. Online protocols prompting users to accept an unknown identity's certificate i) for a single protocol session, ii) forever, or iii) never, offload certificate-authenticity determination to a certificate receiver who cannot possibly make that determination in an informed manner; i.e., the certificate receiver often may not know the real owner of that certificate.

In this paper, we argue that the problem of determining whether a receiver may safely accept attributes (e.g., identity; credentials, such as certificates, groups, roles; and locations such IP addresses, or URLs, physical coordinates) of unknown senders is a special instance of a more general problem of establishing trust in interactive protocols. We define the salient properties of interactive trust protocols and use them to illustrate the usefulness of social collateral in reducing the inherent trust asymmetry in these protocols (Section 2). Then we present a social-collateral model in which receivers are able to accept attributes from *unknown senders* in a safe manner based on explicit recommendations made by social relations; e.g., by their friends, relatives, collaborators (Section 3). We use the notion of the *social collateral* as a measure of both social relations and of "tie strength[1]" among individuals to provide different degrees of accountability for accepting attributes of unknown senders on an *ad hoc* basis (Section 4). Our model is robust in the face of several specific attacks, such as impersonation and tie-strength-amplification attacks (Section 5). The key feature of our model is that a user only needs to perform a single informal measurement of the tie strength between his/her friend and an unknown sender, which is represented by a simple visual diagram. Preliminary experiments with visualization of measured tie strength in a social network indicate that the model is usable by ordinary protocol participants (Section 6).

[1] *Tie strength* is the technical term that refers to the closeness, social proximity, or propinquity of two individuals.

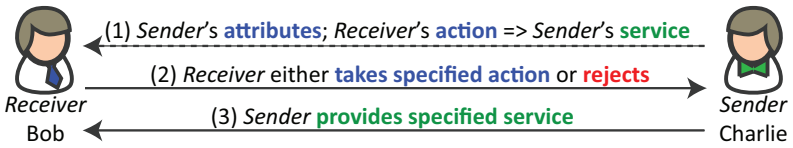

Fig. 1. An interactive trust protocol. When both parties are honest (i.e., complying with the protocol specifications), both are better off after protocol execution.

2 Interactive Trust Protocols

A receiver's decision of whether to accept a sender's attributes is an instance of the more general problem of input trust. In fact, one can show that a receiver's decision to accept input from an unknown sender, where (1) the sender and receiver cooperation benefits both and (2) lack of cooperation benefits the sender and causes the receiver to incur a loss, is an instance of a classic trust problem of behavioral economics [9]. This problem also manifests itself in interactive trust protocols often found in online social applications. The following three generic steps, which are illustrated in Figure 1, characterize these protocols:

1. A sender invites a receiver to participate in a social protocol. (The invitation can be implemented by an explicit protocol message sent to a specific receiver, or by an open invitation posted on a website to any receiver). The sender's invitation comprises the sender's attributes and protocol specification; i.e., if the receiver takes a specified action, the sender will provide a service that will benefit the receiver. For example, the action required of the receiver may be to click on a link provided by the sender, disclose personal information (e.g., personal identification, bank account, credit card number), pay for a forthcoming answer to a query or a solution to a problem, or invest in a specified enterprise. The invitation message itself is assumed to be a benign input to the receiver; e.g., it can be verified to be free of malware.
2. The receiver verifies the unknown sender's attributes and follows the protocol specification: he either takes the specified action or rejects. If the receiver rejects, the protocol ends.
3. The protocol specification allows the sender to verify whether the receiver took the specified action. The sender performs the verification and if the verification passes, the sender follows the protocol and provides the specified service. Otherwise, the sender terminates the protocol. The sender can always give the receiver another chance in the future.
4. If the receiver determines that the sender is non-compliant, it never runs this protocol with the sender again. Non-compliance may manifest itself whenever the unknown sender uses spoofed identity, credentials, or location; and provides corrupt service (e.g., incorrect results, messages containing malware) or no service at all.

Interactive trust protocols have three properties: a value promise, asymmetric trust, and expected execution safety.

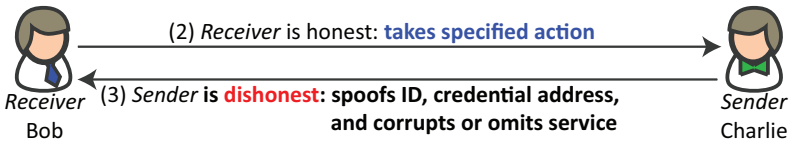

Fig. 2. Asymmetry of an interaction trust protocol. When the sender is dishonest, the sender is better off and the honest receiver is worse off after protocol execution.

Value promise. If both parties are honest, namely they both follow the protocol specification, then both parties are better off after protocol execution. Clearly, the protocol specification must imply that both the sender and the receiver would derive some positive value; i.e., there must be a net benefit in executing the protocol honestly to both parties. We denote the value derived by receiver from honest protocol execution by $HV_R > 0$, and by the sender by $HV_S > 0$. If a positive value does not materialize for either the sender or receiver, the protocol would never be advertised or executed [9].

Asymmetric trust. To obtain the value promised, the sender can always verify and never needs to trust a receiver's honesty. In contrast, a receiver can never verify the sender's honesty before the protocol ends and hence must trust that the sender is honest to obtain the value promised.

The asymmetric trust property implies that the protocol has *asymmetric outcomes* whenever participants are dishonest. If the sender is dishonest, he is better off than when he is honest, since he receives the value promised without having to deliver any service (i.e., any value) to an honest receiver. Specifically, if we denote the sender's positive benefit from dishonesty (i.e., from protocol non-compliance) by DV_S, then the dishonest sender's ill-gotten gain is $Gain_S = DV_S - HV_S > 0$. This means that the sender has an incentive to be dishonest. Furthermore, an honest receiver is worse off whenever the sender is dishonest, since he has to deliver the value promised without obtaining anything in return. Hence, the protocol implies that $Gain_S > 0 \Rightarrow HV_R < 0$. However, if the receiver is dishonest, and claims to have taken the action required by the protocol without actually doing so, the sender terminates the protocol. (Recall that the sender can always discover whether the receiver took the required action.) In this case, $DV_R = 0$ and $HV_S = 0$. Hence, $Gain_R = DV_R - HV_R = -HV_R < 0$, which implies that the receiver has no incentive to be dishonest.

Note that an honest receiver might not even be able to discern a sender's dishonesty until far beyond the end of the protocol. For example, the service provided by a dishonest sender may comprise an arbitrary program whose output behavior cannot always be verified by the receiver. (Verifiability of the output behavior of an arbitrary program is undecidable.) Or, the verification cost may exceed the value of the sender's service to the receiver; e.g., the receiver may have to verify the solution to a co-NP complete problem, which is very unlikely to be possible in polynomial time [9]. Figure 2 illustrates this protocol state.

The trust and outcome asymmetry in interactive trust protocols is inherent: the honest receiver benefits only if the receiver trusts the sender in all protocol runs, whereas the sender does not have to trust the receiver to benefit in any protocol run. Of course, one could modify the protocol specification so that the balance shifts in the favor of the receiver at the expense of the sender, but the inherent trust and outcome asymmetry that characterize these protocols cannot be eliminated. The reasons for this are apparent.

First, in an interactive trust protocol, the *receiver cannot isolate himself from sender's* misbehavior. To receive any value, the receiver has to respond to the server's invitation first, and hence the receiver becomes exposed to a sender's misbehavior, which includes no response at all.

Second, the *receiver may be unable to recover* from sender misbehavior after the protocol ends. Recovery may be expensive or impractical, as it may have complex dependencies on other users' actions, which may be impossible to undo such as recovering a leaked secret or private information. More importantly, the receiver may not even be aware that recovery is necessary. As noted above, he might not be able to detect the effect of a sender's corrupt service until long after recovery becomes impractical; e.g., whether the sender's response message contains malware that damages the receiver may not be an efficiently answered question, if at all.

Third, the *receiver may not have any evidence* of the sender's trustworthiness. Of course, the type of evidence needed depends on whether the sender is a computer or a human using a computer. If the service is a computer, checking trustworthiness evidence reduces to an assessment of correctness evidence in a computational setting. Such evidence, however, is hard to come by: very few systems or services exist that have ever been evaluated at high levels of assurance by most accepted criteria, from the Orange Book (1983) to Common Criteria (2011), and none of these are commodity systems available to all users. Only few commercially available systems, all designed for special applications, have been evaluated at high levels of assurance for the past three decades. If a sender is a human, or a human operating a computer, the notion of trustworthiness becomes strictly stronger than that of computational correctness, as it must encompass evidence of trustworthy human behavior, which is much more difficult to obtain and evaluate. Furthermore, trustworthiness evidence is always about past behavior and hence, even if available, it cannot be a guarantee of sender's present or future behavior in an interactive trust protocol.

Fourth, the *receiver may be unable to deter* a sender's misbehavior. Although intuition suggests that deterrence requires punishment, and punishment requires accountability, it is unknown what punishment and accountability are sufficient for deterrence. Even if sufficient punishment becomes available to deter a dishonest sender, such punishment may be impractical because the sender may be located in a different network jurisdiction than the receiver.

Although asymmetry elimination may not be possible without intervention by other external trusted entities (e.g., trusted third parties), asymmetry reduction may be possible in specific protocol instances. Recent research [9] shows

that the four areas of asymmetry reduction mentioned above, namely isolation, recovery, trustworthiness-evidence evaluation, and deterrence represent the closure of countermeasure types a receiver can employ using both computational and behavioral trust. For interactive trust protocols, all that computational trust suggests can be classified as isolation-, recovery-, and correctness-evidence-based countermeasures for receivers. All that behavioral trust suggests is enhanced receiver preferences (i.e., diminished risk and betrayal aversion) and beliefs in the trustworthiness of the sender. Both preferences and beliefs can be enhanced whenever sender's dishonesty triggers sender's punishment; i.e., it seems natural that betrayal aversion can be decreased and belief in trustworthiness increased by punishment that would deter. Similarly, it seems natural that risk aversion can be decreased and trustworthiness increased by assuring feasible recovery from sender's dishonesty.

Safety. The question faced by a receiver is this: given the inherent asymmetry of interactive trust protocols, is it ever safe for a receiver to accept a sender's invitation and take the required protocol action? The answer to the safety question is unequivocal, if somewhat surprising: under well-defined conditions, a receiver can trust a sender despite the inherent asymmetry of the interactive trust protocols. These conditions are:

1. The protocol is repeated indefinitely in the future, and hence the promise of future value exists;
2. The sender is rational, and hence can compute the present value of future honest behavior; and
3. The present value of future honest behavior exceeds the value of the sender's dishonest behavior (i.e., of DV_S). Hence, the sender has no incentive to cheat.

The questions that need to be answered are how the receiver can compute (i) present value of a sender's future behavior, and (ii) the value of the sender's dishonest behavior. To compute the present value, the receiver must know the sender's discount rate, $r > 0$. Informally, since rational users prefer value in hand over future value, they discount the future value. Discounting future value accounts, among other things, for the uncertainty in obtaining it from a business partner, and interest rates. Since an interactive trust protocol is executed in multiple future sessions (by Condition 1), and the sender's discount rate is $r > 0$ (by Condition 2), at each round, t, of the protocol the value obtained by the sender by executing the protocol honestly is:

$$\frac{HV_S}{(1+r)^t} \quad \text{where} \quad t = 0, 1, 2, \ldots$$

Hence, the present value of all future protocol sessions is

$$HV_S + \frac{HV_S}{1+r} + \frac{HV_S}{(1+r)^2} + \frac{HV_S}{(1+r)^3} + \cdots = \frac{HV_S \cdot (1+r)}{r},$$

and Condition 3 above becomes

$$\frac{HV_S \cdot (1+r)}{r} > DV_S \quad \text{or} \quad r < \frac{HV_S}{DV_S - HV_S} = \frac{HV_S}{Gain_S}.$$

Thus, if $r \geq \frac{HV_S}{Gain_S}$, the sender may not be trusted.

Note, however, that a receiver cannot possibly know the value of the sender's precise discount rate, r, except in very general terms, and hence even if the receiver could compute the ratio $\frac{HV_S}{Gain_S}$, he could not figure out whether r is less than $\frac{HV_S}{Gain_S}$. However, the receiver knows that if $\frac{HV_S}{Gain_S} \to 0$, then $r > \frac{HV_S}{Gain_S}$ and the sender cannot be trusted. In this case, $Gain_S \gg HV_S$, which implies that $\frac{DV_S}{HV_S} \gg 2$. This means that the protocol will have very few sessions before the receiver has to end it. Conversely, if it is a priori known that the protocol will only have a few sessions, say 2, then the receiver should not ever start since the sender has all the incentives to cheat. This implies that all interactive trust protocols that have very only few sessions (e.g., one), may in fact be scams, or deception attempts. Similarly, if previously honest senders discover that they have lost the receivers' trust, they have strong incentives to cheat during the (last) session of the protocol.[2]

Now suppose that, $\frac{HV_S}{Gain_S} \to +\infty$ or that $Gain_S \to 0$. In this case, $r < \frac{HV_S}{Gain_S}$ and the rational sender can be trusted since he has no incentive to cheat during any session of the protocol. Hence, protocol asymmetry would eliminated. However, by the arguments presented above, this is ruled out in interactive trust protocols where, by the definition, $Gain_S > 0$.

3 The Role of Social Collateral

Collateral and trusted third parties. One way to ensure that $Gain_S \to 0$ is to modify the protocol and introduce a third party that is trusted by *both* sender and receiver. The role of the trusted third party (TTP) is simple: the TTP computes $Gain_S$, establishes a collateral value that exceeds $Gain_S$, and collects it from the sender before the first protocol session. If the sender does not comply with the protocol in some session, the TTP uses the sender's collateral and compensates the receiver for his losses. This effectively eliminates a sender's incentive to be dishonest and thus the protocol asymmetry. Of course, for a receiver to accept a sender's invitation to engage in the protocol, the receiver's potential loss must be less than the collateral value. In this case, even if a sender cheats, the receiver never loses anything. In short, two conditions must be satisfied to eliminate protocol asymmetry:

- **Sender's deterrence:** TTP Collateral $> Gain_S$; and
- **Receiver's acceptability:** TTP Collateral > Receiver's Loss.

Although this modification of the interactive trust protocols resolves the asymmetry problem, it is impractical for two reasons:

[2] This fact is also consistent with the observation that insiders, who are trusted to provide honest services to their organizations, are likely to attack their own organization when they suspect that they are about to be fired [21, 22].

1. The modification assumes that an external TTP can be found that is trusted by both sender and receiver. This may be challenging and less than satisfactory: the protocol between a sender who deposits collateral and the TTP who received the collateral is an interactive trust protocol itself and so is that between a receiver and the TTP. In effect, by using a TTP, we have simply removed the asymmetry from the original trust protocol and moved it to the sender and receiver protocols with their common TTP. Hence, we have not completely eliminated trust asymmetry.
2. More importantly, the modified trust protocol is unlikely to start whenever the sender invites multiple receivers, since it does not scale: the sender may be unable to post separate collateral for every receiver who might accept the sender's invitation to engage.

Social collateral, deterrence, and acceptability. We now show that it is possible to reduce the asymmetry of an interactive trust protocol between a receiver and an unknown sender *without relying on a TTP* to collect, hold collateral, and compensate the receiver for his loses when needed. Let us assume that a social relation exists between the receiver and a third party who also has a social tie to the unknown sender. That is, the third party may be a close friend, immediate-family member, or colleague of long standing of the receiver. In the social collateral model [15], this implies that the third party has social collateral with the receiver and any misbehavior by the third party would cause loss of the collateral. In particular, the third party would be deterred from providing false inputs to the receiver by the loss of the social collateral. Thus, all recommendations made to the receiver are likely to be correct, or at least not intentionally deceitful. Also, the existence of a social relationship implies that any trust protocol between the receiver and the third party could be repeated indefinitely and the present value of future honest protocol sessions is high. Furthermore, it implies that any trust protocol between the receiver and the third party can always be initiated.

The role of the third party in interactive trust protocols is *not* that of a TTP. First, the third party need not be trusted by *both* the receiver and sender. In fact, the sender and the third party need not trust each other at all. They only need to have a social tie that is sufficiently strong so that the third party's recommendation to the receiver regarding the sender is, in fact, accepted by the receiver. Furthermore, the trust between the receiver and the third party is already fully captured by existing social collateral and need not be established by yet another trust protocol. In short, the trust asymmetry between the receiver and the unknown sender is reduced without requiring a TTP.

The remaining questions are whether (1) the present value of future honest protocol sessions implied by social collateral exceeds the third party's value of dishonest behavior (i.e., false recommendations) in any future protocol session, and (2) the receiver considers the social collateral acceptable. The answer to the first question would determine whether the loss of a third party's social collateral with the receiver is sufficient to deter any misbehavior (i.e., bad recommendation) by the third party. The answer to the second question would determine

whether the third party's social collateral exceeds the receiver's loss resulting from potential misbehavior of the third party. While these questions cannot be answered without taking into account the specifics of an interactive trust protocol, evidence indicates that loss of social collateral has non-negligible deterrent value, and that the reduction of asymmetry between the third party and receiver has a direct relationship to loss exposure by the receiver [15]. Hence, in our model for attribute authentication we rely on the following hypothesis and asymmetry-reduction criterion.

Deterrence Hypothesis: *The loss of a social relation deters misbehavior.*[3]
Asymmetry-Reduction Criterion: *The greater a receiver's exposure to loss is, the more social collateral is required.*

4 A Social Collateral Model for Attribute Authentication

All characteristics of an interactive trust protocol are found in the online social network problem of accepting an invitation from an unknown sender. In online social networks, the receiver can materialize the *value promise* only by accepting the sender's attributes, even when attribute authentication may be impractical in the absence of an identification and authentication infrastructure. For example, when the receiver accepts the sender's attributes, the receiver's potential benefits are as follows: (1) build new social, professional, or business connections with the sender and his friends; (2) use sender's services with the assurance that the sender is accountable, since his identity and social connections are known; and (3) develop his/her own social network connections by building future strong ties with the sender and be recommend by the sender to others. *Asymmetry* is also evident: the receiver knows nothing about the sender attributes' authenticity whereas the sender knows everything about them. The *safety problem* also arises here: is it ever safe to accept the attributes of an unknown sender in the absence of an identification, authentication, and accountability infrastructure?

A receiver's decision to accept a third-party's authentication of an unknown sender's attributes presented in Section 1 can now be framed as a trust decision to be made in an interactive protocol where a third party has (1) a *social relation* with the receiver and (2) a *social tie* with the invitation sender and the receiver. This scenario is illustrated in Figure 3, where $SC(A)@B$ denotes the social collateral which third party A *has* with receiver B as the result of their friendship, whereas $SC(C)@A$ denotes the social collateral *assigned* by receiver B to the signed recommendation made by third party A for an attribute of the unknown sender C.

Social ties. Unlike social relations, which imply the existence of social collateral, we use social ties only as a measure of the *social distance* between two parties. Although they do not necessarily imply existence of collateral, social ties serve

[3] Recent evidence shows that loss of social relations deters more than the law, even when both law and loss of social collateral fail to provide sufficient deterrence for specific forms of misbehavior (e.g., insider misuse of permissions) [14].

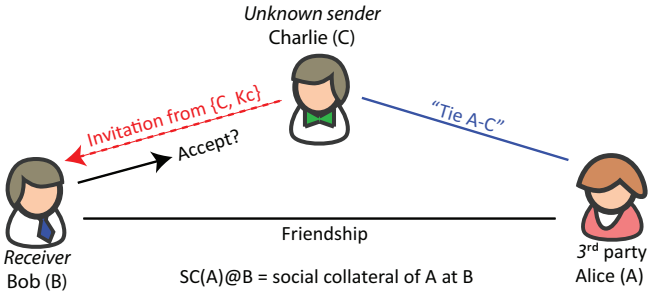

Fig. 3. An interactive trust protocol with a social relation. In this figure, 3^{rd} party A has a social relation with receiver B and a social tie with sender C. K_C stands for the public key of C.

as an indication of the knowledge one party has about the other. Stronger ties imply more accurate knowledge, and this in turn serves as the basis for more credible recommendations. Hence, being able to measure the strength of a tie between receiver B's friend A and unknown sender C in a manner that can be easily evaluated by B becomes important, particularly since our model requires B to assign a social collateral value to the strength of a tie between A and C (discussed below).

Social science research has studied a variety of parameters that capture the strength of ties between individuals. Gilbert and Karahalios [6] have recently showed that four relatively simple parameters are sufficient for determining tie strength in practice: communication reciprocity [5, 11, 19], existence of at least one mutual friend [24], recency of communication [20], and interaction frequency [7, 11]. Our model relies on the ease of measurement, display, and understanding of these parameters by humans since it requires assessment of tie strength values and assignment of social collateral to them. In addition to these four parameters, we use length of the relationship as an additional tie-strength indicator. We do this because the length of a relationship increases accountability by adding a significant degree of moral responsibility to reporting authentic attributes of unknown senders. Shneiderman's work on the rich feedback about content quality provided by patterns of past performance online [25] supports the inclusion of this additional parameter.

The tie strength parameters are collected from a variety of online sources; e.g., online social networks, email, peer-to-peer (P2P) communication, physical-encounter evidence provided by GPS-enabled phones, accounts of phone communications. Some of these (required) parameters could be deliberately manipulated by a single individual; e.g., communication recency may be inflated by spurious emails and P2P messages. However, not *all* parameters can be manipulated *simultaneously* unilaterally to generate consistent false tie-strength measurements, since not all parameters are under the control of a single individual. For example, physical encounters, accounts of reciprocity in phone calls require both individuals to act. Nevertheless these parameters could be *artificially inflated*

by collusion between two individuals. Furthermore, some parameters under user control could be *decreased* whenever individuals collude to hide the strength of their social tie. Hiding the strength of a social tie may not necessarily be a malicious act designed to misinform an unsuspecting receiver.

Privacy Concerns. While the privacy of his tie to the unknown sender may be less of a concern for the third party with respect to his friend the receiver, revealing the *strength* of his social tie with the third party may violate the privacy concerns of the unknown sender. However, this is not a surprise: very often, protocols that establish authenticity conflict with privacy in an unavoidable manner [16]. However, in interactive trust protocols, the potential loss of privacy is under the control of the parties who are affected by it. That is, the sender and the third party can decide how much, if any, of their social tie strength to reveal to a receiver. Furthermore, this decision can be unilaterally taken or negotiated; e.g., the third party may refuse to sign the strong tie evidence requested by a sender, and the sender may selectively remove or decrease the values of some revealing parameters under individual control. However, not all parameter values can be simultaneously decreased, as some values may be provided by network services outside individual user control; e.g., phone and e-mail account information. The use of these parameters is required by the receiver so that he could assign social collateral to the social tie in an reasonably accurate manner for deterrence purposes.

Social collateral assignment. In our model, a recommendation for the authenticity of an unknown sender's attributes comprises (1) the specification of the attribute whose authenticity is vouched by the third-party recommender, (2) the evidence of the social tie between the unknown sender and recommender, and (3) the recommender's signature. In contrast with $SC(A)@B$, which is a direct measure of the friendship between A and B, to assign collateral value $SC(C)@A$ to third party A's recommendation, receiver B verifies A's signature, using A's public key which we assume B already has, and evaluates the social tie evidence included in the recommendation.

In assigning social collateral to the third party A's recommendation for unknown sender C's attributes, higher collateral values correspond to stronger evidence of the social tie between C and the recommender A. Since some of C's attributes may require more knowledge about C for authentication, a stronger tie between A and C becomes necessary. This is the case because receiver B's risk of security exposure caused by accepting a false attribute as authentic for a particular application may be higher for some attributes than for others. Hence, that risk must be offset by recommender A's better knowledge of, and stronger tie to, the unknown third party C. For example, accepting C's identity as authentic would require lower tie strength than accepting C's public key, since the public key may be used to set up a secure channel for the later transmission of sensitive data, whereas C's identity may be used merely for granting C read access to low-sensitivity objects. Similarly, accepting a set of attributes would require higher tie strength than accepting a proper subset of those attributes.

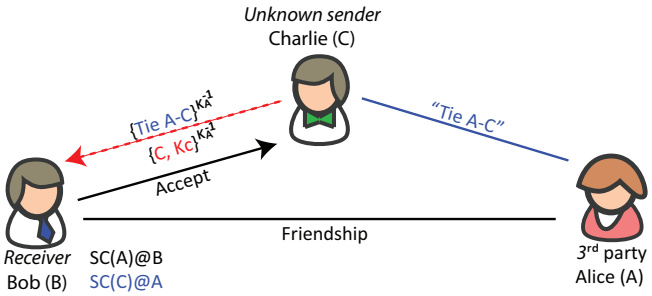

Fig. 4. Accepting a certificate from an unknown sender signed by a friend.

Acceptability and deterrence. The above discussion indicates that receiver B has a particular collateral threshold for accepting an attribute of an unknown sender C in his application. Let $T_B(app, attr)$ denote that threshold. Hence, receiver B verifies that

$$Acceptability: \quad SC(C)@A \geq T_B(app, attr)$$

where $T_B(app, attr) \geq 0$ is a measure of the loss incurred by B's application, *app*, if attribute, *attr*, is unauthentic.

In our model, recommender A looses her social collateral with friend B, namely $SC(A)@B$, if the recommendation to accept an attribute, *attr*, as authentic in B's application, *app*, turns out to be false. Our deterrence hypothesis suggests that loss of this social collateral would prevent A from making false recommendations to B. However, A faces a clear case of moral hazard. That is, if the social tie between recommender A and the unknown sender C is stronger than the friendship between A and receiver B, C could conceivably bribe A to make a false recommendation to B. This fact has been pointed out in the social collateral model of Karlan et al. [15]. Hence, B has to verify that his friendship with A is stronger than A's social tie to C.

$$Deterrence: \quad SC(A)@B - SC(C)@A \geq P_B(app, attr)$$

where $P_B(app, attr) \geq 0$ is a measure of the net loss of social collateral incurred by A if A's recommendation attribute, *attr*, for B's application, *app*, is unauthentic.

Figure 4 illustrates B's acceptance of a public-key certificate recommendation for unknown sender C from his friend A.

Second independent opinion. Suppose that receiver B's acceptability check for unknown sender C's attribute, *attr*, does not pass for application, *app*, because the tie strength between A and C is too low. To ensure that his rejection of C's invitation is justified, B can seek a second, independent third-party's recommendation. To do so, B searches sender C's social graph to determine whether C has a social tie with any other of B's friends. If this search returns a non-empty

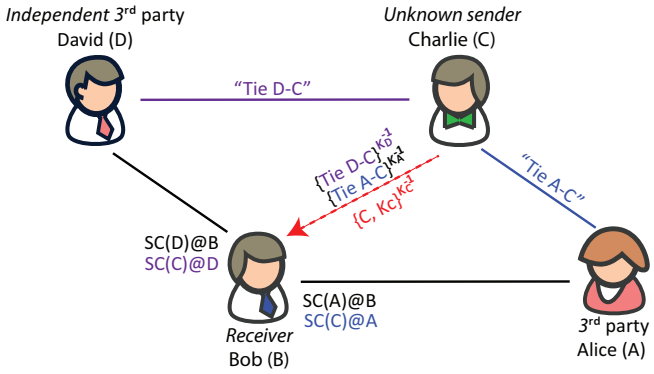

Fig. 5. Accepting a self-signed certificate from an unknown sender C. Receiver B seeks second opinion from 3^{rd} party friend D who is independent from friend A.

list of B's friends who have a social tie with C, then B selects, at random, another possible recommender for C's attribute, from the list, say friend D. Then B informs C of the need to obtain a signed recommendation from D and provides it along with C's invitation. When this second recommendation is made available, receiver B's acceptability and deterrence checks become:

$$\begin{aligned} &\textit{Acceptability}: SC(C)@A + SC(C)@D \geq T_B(app, attr) \\ &\textit{Deterrence}: \quad SC(A)@B - SC(C)@A \geq P_B(app, attr) \textit{ and} \\ &\qquad\qquad\quad SC(D)@B - SC(C)@D \geq P_B(app, attr) \end{aligned}$$

Note that, in the social collateral model of Karlan et al. [15], the social collateral available on two separate paths between a source and destination is unconditionally additive. In contrast, in our model, additivity is conditioned on B's random selection of a second recommender friend, D, who has a social tie with unknown sender C. The random choice of D implies that C's chances of bribing both of B's independent friends, A and D, to deliberately reduce their tie strength evidence simply to pass B's deterrence checks and vouch for unauthentic C attributes, are significantly diminished.

Figure 5 illustrates B's acceptance of a public-key certificate for unknown sender C based on the independent recommendations of his friends A and D.

Forwarded recommendations. Suppose that a social tie between unknown sender C and any one of receiver B's friends does not exist. Instead, a social tie between C and a friend of A, namely E, exists. This case is illustrated in Figure 6. Furthermore, suppose that B's friend A has accepted a recommendation from her friend E regarding the authenticity of unknown sender C's attribute (i.e., public key certificate $\{C, K_C\}^{K_E^{-1}}$, and that A is willing to forward E's recommendation to B along with E's public key.

In this case, unknown sender C can present two pieces of evidence to receiver B to justify the authenticity of C's attribute: i.e., certificate $\{C, K_C\}^{K_E^{-1}}$. The

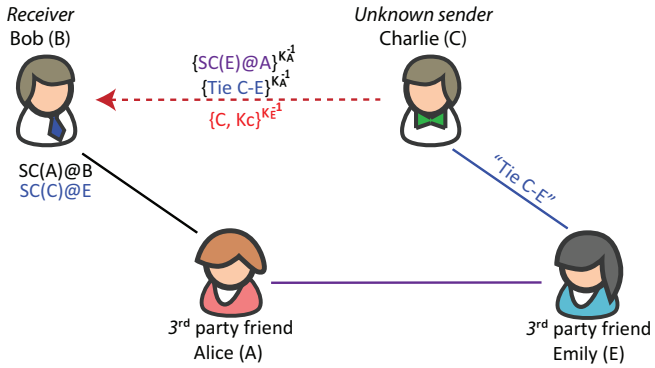

Fig. 6. Accepting an unknown sender C's attribute signed by a friend (E) of a friend (A).

first is the evidence of E's tie to C signed by A, namely $\{Tie(C-E)\}^{K_A^{-1}}$. The second is A's assessment of E's social collateral at A (i.e., E's friendship with A) signed by A, namely $\{SC(E)@A\}^{K_A^{-1}}$. Given these pieces of evidence, should receiver B accept the forwarded authentication of C's attribute, $\{C, K_C\}^{K_E^{-1}}$? To accept, B has to evaluate both the recommendation itself and whether A's forwarding of E's recommendation is warranted by B's social relation (friendship) with A.

To evaluate E's recommendation of C's attribute, B asks the following question: are A's deterrence and acceptability criteria for E's recommendation at least as strong as mine (B's)? To answer this question, B has to apply the two criteria using his own parameters, namely $T_B(app, \{C, K_C\})$, $P_B(app, \{C, K_C\})$, and his assignment of social collateral to the tie $\{Tie(C-E)\}$, namely $SC(C)@E$, after verifying signatures appropriately. B's criteria are:

$$\begin{aligned} &Acceptability: SC(C)@E \geq T_B(app, \{C, K_C\}) \\ &Deterrence: \quad SC(E)@A - SC(C)@E \geq P_B(app, \{C, K_C\}) \end{aligned}$$

To determine whether A's forwarding of E's recommendation is warranted, receiver B again applies his criteria to friend A's and social tie between A and E. To do so B uses A's assignment of social collateral to her friendship with E, namely SC(E)@A, as follows:

$$\begin{aligned} &Acceptability: SC(E)@A \geq T_B(app, \{C, K_C\}) \\ &Deterrence: \quad SC(A)@B - SC(E)@A \geq P_B(app, \{C, K_C\}) \end{aligned}$$

In short, when these two pieces of evidence are made available to receiver B by unknown sender C, receiver B's acceptability and deterrence checks become:

$$\begin{aligned} &Acceptability: min\{SC(E)@A, SC(C)@E\} \geq T_B(app, \{C, K_C\}) \\ &Deterrence: \quad min\{SC(A)@B - SC(E)@A, SC(E)@A - SC(C)@E\} \\ &\qquad\qquad \geq P_B(app, \{C, K_C\}) \end{aligned}$$

We note that these two checks are applicable to other attributes of C, not just certificate $\{C, K_C\}$.

5 Model Robustness

The social collateral model defined above assumes two types of adversarial attacks. In the first type, the adversary is the unknown sender C who attempts to impersonate a false identity or provide a false certificate for a known identity to an unsuspecting receiver B. Adversary C *unilaterally manipulates* tie strength parameters in an attempt to increase his chances of successfully convincing B to accept his unauthentic attributes. In effect, adversary C may try to inflate a recommendation from B's friend A. However, this type of attack cannot succeed in the protocols proposed above for three complementary reasons. First, recommender A will not agree to endorse (i.e., sign) inflated tie strength parameters since she is deterred by the social collateral loss with receiver B. Second, the inflation of all social tie parameters will fail because, as discussed in the previous section, some parameters may not be controllable by the user and others may require collusion with the recommender. Third, our model offers no incentive to recommender A to collude with sender C and endorse tie strength parameters inflated by C. The deterrence check performed by B would reject very strong ties between C and A. Hence, the moral hazard to which A might be exposed by refusing to endorse C's inflated social tie parameters (e.g., the distrust revelation problem [17]) does not materialize in our model.

In the second type of adversarial attack, the unknown sender C colludes with receiver B's friend A to *conceal and diminish* the real strength of their tie and induce B to accept C's false credentials. (Recall that collusion between A and C to inflate their tie strength is countered by B's deterrence check, as discussed above.) A particular instance of this attack may materialize when unknown sender C is in fact a secret Sybil of B's friend A. Our model addresses this type of attack in three distinct ways.

Independent second opinion. If the social-tie strength is too low and does not pass receiver B's acceptability threshold, B would automatically request an independent second opinion from friend D (viz., in Figure 5). Should a second independent opinion not be available, B's acceptability test would fail. It is rather unlikely that C could anticipate and "bribe" a randomly chosen independent provider by a second opinion. Furthermore, note that the independence of the second opinion could not be easily manipulated by sender C since the choice of the second-opinion provider, D, is exclusively receiver B's. Sender C has no say in it. Furthermore, unknown sender C has no clue of the amount of collateral receiver B's friends A and D have with B. All C sees in the social graph are "friend" connections. C would have to bribe *all* of B's "friends" he knows since he does not know B's future second-opinion provider. Thus, it seems unlikely that *both* Bob's friends A and D would collude with their acquaintance C against receiver B. Also note that whether A and D are (not) connected on the social graph is irrelevant to our notion of recommendation independence.

Deterrence against threshold probing. Receiver B can also detect whether recommender friend A and unknown sender C probe his acceptability threshold in a particular application by repeatedly including different tie-strength parameter values. First, repeated recommendations for the same identity C issued by A within a given time interval would automatically cause a second-opinion request by B. Second, repeated recommendations for different identities corresponding to real C, would require that all identities be related to B's other friends since a second opinion request would fail otherwise. Furthermore, low tie strength evidence for any of C's identities would have to be maintained by A for all C's recommendations in the social network over an interval of time. Otherwise, B's false recommendations of C could be detected. Third, a couple of failed recommendation attempts by A would cause A to lose his collateral with B.

Mandated tie strength parameters. If the colluding parties, namely A and C, could control *all* tie strength parameters simultaneously, they might still be able to discover a (narrow) range of parameter values that are both acceptable to B and pass his deterrence check. Our model also addresses this type of attack that not all recording mechanisms for measurable tie strength parameters be under a user's control. Hence, not all parameters could be decreased simultaneously to create consistent false evidence of weak social ties. We require that some parameters that could not be manipulated by any colluding parties be used in all recommendations. These parameters would be routinely provided and endorsed by third parties such as phone, e-mail, and other service providers.

6 Usability

We have conducted an extensive set of user studies to verify real users' ability to display, understand, and evaluate the tie strength between parties. We stress that this is the only parameter of our model that needs to be explicitly measured and displayed to users. In contrast, the collateral values of social relations of a user can be reasonably accurately estimated by the user himself/herself. Our usability studies are reported in detail elsewhere [18]. In this section we summarize those findings.

6.1 Visualization of "Tie Strength" Evidence

Tie strength can be visualized to help authenticate online identities. For example, based on the social relationship as depicted in Figure 5, receiver Bob can decide to accept an online "friend" invitation from unknown sender Charlie as follows: if the invitation contains visual tie strength evidence endorsed by their mutual friends (Alice and David in this case), Bob accepts Charlie's invitation based on the social collateral he assigns to the tie strength between Charlie and his two mutual friends.

We have formulated the details of visualizing tie strength, and Figure 7 is an example of tie strength visualization. This visualization displays six parameters:

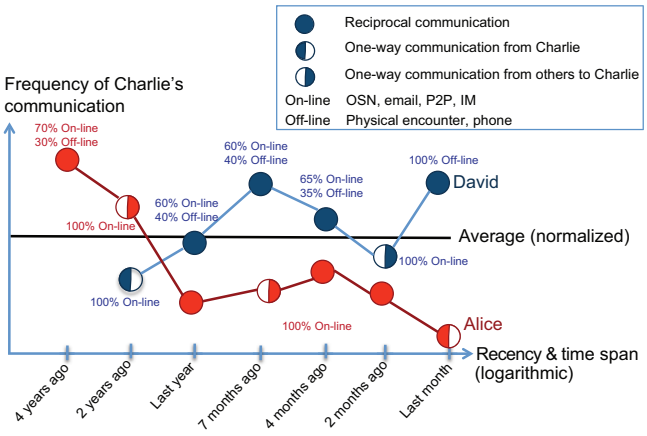

Fig. 7. Visualization of tie strength between Charlie and Bob's friends (Alice and David). This diagram displays how frequently Charlie has been interacting with Alice (red graph) and David (blue graph).

- **Frequency of communication:** This parameter is represented on the y-axis. Bob can compare how frequently Charlie has interacted with Alice by comparing Alice's graph with the *average* interaction frequency line which represents the interaction frequency that Bob had with all his other friends (i.e., 4 years ago, Charlie interacted more frequently with Alice than with all his other friends).
- **Length of relationship:** As represented on the x-axis, this parameter shows how long Charlie has known Alice and David.
- **Reciprocity of communication:** The variations of coloring schemes on a circle represent the reciprocity information. A fully-colored circle represents that two people communicate reciprocally, and a half-colored circle represents one-way communication where one party attempts to interact but the other party does not respond.
- **Selected mutual friends:** The individual graphs in visualization correspond to selected mutual friends between the invitation sender and the receiver. Figure 7 displays Alice and David as Bob's selected mutual friends.
- **Recency of interaction:** The rightmost point on the graph represents how recent the interaction was between the sender and the mutual friend. In Figure 7, Charlie's most recent interaction with David was last month.
- **Communication type:** People can communicate using (1) on-line channels (e.g., Online Social Networks (OSNs), emails, Instant Messangers), or (2) off-line channels (e.g., physical encounter, phone conversations). Labeling the communication type empowers the receiver to judge the approximate strength of ties. For example, a "100% on-line" label for the entire length of relationship may indicate individuals who have only established a relationship through purely on-line means.

6.2 Usability Evaluation: a Facebook Application

In order to test whether the tie strength visualization help users make correct authentication decisions, we have developed a Facebook application that plots interaction frequencies. After querying the Facebook database according to the user's policy, this application retrieves a stream of wall posts between the sender and mutual friends, and plots interaction frequencies on a graph.

Procedure. We designed an online user study and recruited 93 participants using Amazon Mechanical Turk. All our participants were living in the U.S.

We asked each participant to download our Facebook application and run it to check the tie strength visualization of their pending Facebook invitation senders or their friends (if they did not have any pending invitations). We asked each participant to run the application at least 3 times.

We then asked them to provide feedback on our tie strength visualization. More specifically, we asked questions related to (1) how understandable is the visualization, (2) how easy is the application to use, and (3) whether they would accept an invitation even if this application displays below-average interaction frequency with the participant's friend(s).

Results. Overall, participants provided promising feedback.

- **Understandability:** 85% of the participants indicated that they understood tie strength of people as shown on the visualization, and 85% indicated that Figure 7 was a good way of displaying tie strength.
- **Robustness:** 90% indicated that they would not accept an invitation if the graph were placed below the average interaction frequency.
- **Usability:** 83% indicated that the visualization was easy to use, and 88% mentioned that our authentication application was easy to use. When we asked for possible future use, 84% expressed likeliness to use our application before accepting invitations.

The study results confirm that providing tie strength visualization to users is a promising direction to help users authenticate online identities.

7 Conclusions and Future Work

This paper introduces the notion of interactive trust protocols and illustrates how to establish attribute-authentication trust between two untrusting parties (e.g., between service providers and service receivers) who do not share a common trusted third party (TTP) or trust infrastructure (e.g., a public-key infrastructure). While helpful in many cases, TTPs create additional complexity and uncertainty, and sometimes become an attractive attack target. More fundamentally, the need for TTPs would beg the very question we want to answer, namely how can we establish trust between two previously untrusting parties. To remove the need for a TTP, we used a social collateral model inspired by those found in behavioral economics [15].

Interactive trust protocols could also be used in modeling the basic steps of online scams and/or deceptions. Most such protocols include a value proposition, asymmetric trust/outcomes, and an unsatisfied, or unsatisfiable, safety condition. Hence, the use of these protocols to model scams and deceptions would offer a way to detect patterns of possible scams and alert unsuspecting users. A good starting point would be to use these protocols for the few real-life cases of scams described by Stajano and Wilson [26].

Another extension of this work would be to develop social collateral models for *trust networks* [8], perhaps along the lines of those studied in behavioral economics. For example, the modeling of "agency problems" using Greif's well-known model [12] may in fact work better in 21^{st} century's computer networks than in 11^{th} century's coalitions of Maghribi traders [4, 10].

8 Acknowledgments

This research was supported in part by NSF under awards CCF-0424422 and CNS-1050224. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of CMU, NSF or the U.S. Government or any of its agencies.

References

1. Sophos Facebook ID Probe. `http://www.sophos.com/pressoffice/news/articles/2007/08/facebook.html`.
2. L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks. In *Proceedings of WWW*, 2009.
3. Economist. Duly notarised. `http://www.economist.com/blogs/babbage/2011/09/internet-security`, Sept. 2011.
4. J. Edwards and S. Ogilvie. Contract Enforcement, Institutions and Social Capital: the Maghribi Traders Reappraised. *CSEIFO Working Paper*, March 2008.
5. N. E. Friedkin. A Test of Structural Features of Granovetter's Strength of Weak Ties Theory. *Social Networks*, 1980.
6. E. Gilbert and K. Karahalios. Predicting Tie Strength With Social Media. In *Proceedings of the* 27^{th} *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2009.
7. E. Gilbert, K. Karahalios, and C. Sandvig. The Network in the Garden: An Empirical Analysis of Social Media in Rural Life. In *Proceedings of the* 26^{th} *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2008.
8. V. Gligor, A. Perrig, and J. Zhao. Brief Encounters with a Randomkey Graph. In *Proceedings of the* 17^{th} *Security Protocols Workshop*, April 2009.
9. V. Gligor and J. M. Wing. Towards a Theory of Trust in Networks of Humans and Computers. In *Proceedings of the* 19^{th} *International Workshop on Security Protocols*, March 2011.

10. J. Goldberg. Making reputation work: re-examining law, labor and enforcement among Geniza businessmen. Before and Beyond Europe: Economic Change in Historical Perspective (Yale University), February 2011.
11. M. S. Granovetter. The Strength of Weak Ties. *The American Journal of Socialogy*, 1973.
12. A. Grief. Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders Coalition. *American Economic Review*, June 1993.
13. N. Hamiel and S. Moyer. Satan Is On My Friends List: Attacking Social Networks. In *Black Hat Conference*, 2008.
14. Q. Hu, Z. Xu, T. Dinev, and H. Ling. Does Deterrence Work in Reducing Information Security Policy Abuse by Employees? *Communications of The ACM*, 2011.
15. D. Karlan, M. Mobius, T. Rosenblat, and A. Szeidl. Trust and Social Collateral. *The Quarterly Journal of Economics*, August 2009.
16. S. T. Kent and L. I. Millett, editors. *Who Goes There? Authentication Through the Lens of Privacy*. National Academies Press, 2003.
17. T. H.-J. Kim, L. Bauer, J. Newsome, A. Perrig, and J. Walker. Challenges in access right assignment for secure home networks. In *Proceedings of the* 5^{th} *USENIX Workshop on Hot Topics in Security (HotSec '10)*, 2010.
18. T. H.-J. Kim, A. Yamada, V. Gligor, J. I. Hong, and A. Perrig. RelationGrams: Tie-Strength Visualization for User-Controlled Online Identity Authentication. Technical Report CMU-CyLab-11-014, Carnegie Mellon University, 2011.
19. D. Krackhardt. The Strength of Strong Ties: The Importance of *Philos* in Organizations. *N. Nohria and R. Eccles (eds.), Networks and Organizations: Structure, Form,and Action*, 1992.
20. N. Lin, P. W. Dayton, and P. Greenwald. Analyizing the Instrumental Use of Relations in the Context of Social Structure. *Sociological Methods Research*.
21. A. P. Moore, D. M. Cappelli, T. C. Caron, E. Shaw, D. Spooner, and R. F. Trzeciak. A Preliminary Model of Insider Theft of Intellectual Property. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 2011.
22. A. P. Moore, D. M. Cappelli, and R. F. Trzeciak. The "Big Picture" of Insider IT Sabotage Across U.S. Critical Infrastructures. Technical Report CMU/SEI-2008-TR-009, Carnegie Mellon University, 2008.
23. T. Ryan. Getting in Bed with Robin Sage. In *Black Hat Conference*, 2010.
24. X. Shi, L. A. Adamic, and M. J. Strauss. Networks of Strong Ties. *Physica A: Statistical Mechanics and its Applications*.
25. B. Shneiderman. Designing Trust into Online Experiences. *Communications of the ACM*, 2000.
26. F. Stajano and P. Wilson. Understanding Scam Victims: Seven Principles for Systems Security. *Communications of the ACM*, 2011.