Realistic Threats to Self-Enforcing Privacy

Giampaolo Bella [1,2], Francesco Librizzi[1] and Salvatore Riccobene [1]

[1]Dipartimento di Matematica e Informatica, Università di Catania
Viale A.Doria, 6 — I-95125 Catania, ITALY

[2]SAP Research Labs, 805 Avenue du Dr Maurice Donat, 06250 Mougins, France

{giamp,librizzi,sriccobene}@dmi.unict.it

Abstract

A recent privacy protocol for secure e-polls aims at ensuring the submitting individuals that the pollster will preserve the privacy of their submitted preferences. Otherwise the individuals can indict the pollster, provided that the pollster participates actively in this phase. The analysis of the protocol in a realistic threat model denounces that a malicious pollster that abuses the private preferences by disclosure will arguably not help out during its own indictment. Therefore, the protocol ensures insufficient fairness among their participants because it gives the pollster some advantage over the individuals. Two variant protocols are introduced and analysed in the same threat model — one is found to move the advantage over the individuals, the other is found to achieve a satisfactory level of fairness.

Keywords: *network privacy, electronic polling, threat model*

1 Introduction

The constantly increasing variety of services available from computer networks such as the Internet is no longer surprising in the current decade. The individuals' privacy is becoming a major concern. This paper focuses on a particular networked service: electronic polling systems (e-polling in brief). E-polling is rather different from the perhaps better known e-voting. While the results of an electronic poll typically are statistics expressing individuals' preferences, those of an electronic election must be precise counters. Moreover, it is not uncommon that an individual submits false responses to a poll, but the statistics will most probably absorb the fakes.

Statistical investigations are required about a large variety of topics, but typically seek individuals' private information such as the use of drugs or alcohol. It is clear that the submitters care about the privacy of their data, and consequently the role of the pollster is delicate. The pollster must not publish the raw data otherwise it would commit a privacy breach. However, it must be allowed to use them to compute some statistics that will have to be inspected by other parties or eventually published.

A secure e-polling protocol should ensure data privacy even against a malicious pollster. For example, the pollster might publish the private data rather than the statistics based on them. The individuals would then be able to detect the privacy breach but would have no means to demonstrate that it was exactly the pollster the entity that published their data.

Golle et al. recently proposed a simple technique towards privacy-preserving e-polling [7]. It prescribes the individuals to submit their preferences bundled with some "baits". Intuitively, these will serve to indict a malicious pollster because it is unable to remove the baits from the submitted data. Therefore, even with malicious aims, the pollster can only publish the submissions as they arrive, and hence both real data and baits. More precisely, the baits are bits of information that will not significantly affect the statistics that the pollster is meant to make of the submissions.

As the protocol by Golle et al. aims at self-enforcing privacy, we name it SEP for simplicity. The main goal of our research is to analyse SEP against current real-world threat models. It is understood that the conclusions of a security analysis are strongly dependent on the threat model that is adopted. One of our finding proceeds from the observation that the pollster is required to participate actively in the indictment that the individuals may start against itself. Our analysis of this phase concludes that a malicious pollster that published the raw private data arguably will not want to participate in its own indictment.

This finding indicates that the pollster has some advan-

tage over the individuals when it is indicted. In this light, we advance two variants of the original protocol. One is found to move the advantage from the pollster onto the individuals, and therefore is unsatisfactory. The other one, named SEP+, achieves an appropriate level of fairness despite its simplicity — SEP+ is fairer than SEP to its participants in case of indictment.

The organization of this manuscript is simple. The original SEP protocol is described (§2) and then analysed under realistic threats (§3). After that, our variant protocols are defined and analysed under the same threats (§4). Finally come some conclusions (§5).

In the description we use the word 'principal' to address the individual who submits information to the pollster.

2 The Self Enforcing Privacy Protocol

The key ideas underlying all techniques proposed by Golle et al. [7] are rather simple, as depicted in Figure 1. To trace data after their transmission, each individual is required to add to her preferences P_1, P_2, ..., P_{p_1} some information that must link the preferences with the pollster. This additional information B_1, B_2, ..., B_{b_1} serves as baits. Therefore, each individual in fact transmits a bundle containing her preferences and the baits.

In particular, it is important that the baits do not compromise the results of the statistical investigation, and the pollster be unable to distinguish if data are preferences or baits. Thus, if the pollster is dishonest and publishes or sells individuals' personal data, then the individuals must be able to indict it publicly. The indictment is possible exactly because the published data contain individuals' baits. It is also necessary to ensure that the pollster cannot be indicted illicitly. Since the individuals have the baits, they could insert them in a fake personal data collection, then publish the collection, and finally indict an honest pollster. A fair scheme must make that indictment impossible.

The SEP protocol implements the previous scheme fairly. It adopts the RSA encryption scheme [6, 8] to implement the baits. The main idea is that each individual computes the baits using a hash function that the pollster makes public. That function is required to have as image the set of ciphertexts that can be produced using the underlying cryptosystem. As a result, the pollster will be unable to discern whether a ciphertext was produced using the hash function, in which case it is a bait, or using the actual encryption algorithm, in which case it is a real preference.

We can now move on to describe the protocol in detail. It is composed of four main phases plus the indictment phase.

- ***Setup.*** The pollster publishes the parameters for the encryption algorithm E and two hash functions. One, named h, is a standard hash such as SHA-256; the other one, named g, is a special hash function whose image is the same as the encryption function's, i.e. $\mathcal{I}m(g) = \mathcal{I}m(E)$.

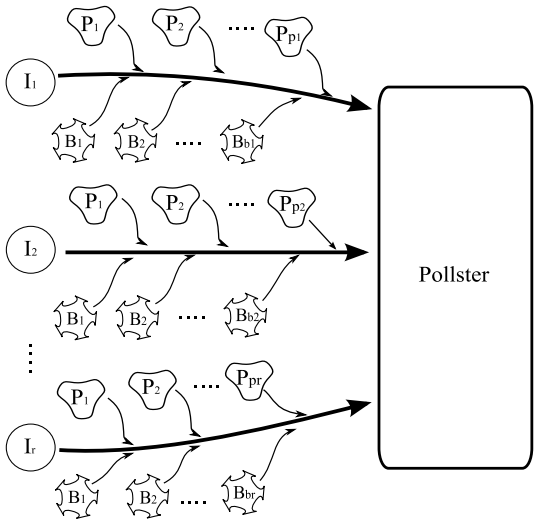

Figure 1. The original SEP protocol

- ***Sending a bit to the pollster.*** The individual's preference is sent bit by bit. To send a bit $b \in \{0, 1\}$ the individual I_i chooses a random value r such that the least significant bit of $h(I_i \parallel r)$ is b. The individual sends $\langle I_i, E(r)\rangle$ to the pollster.

- ***Sending a bait to the pollster.*** To send a bait to the pollster, the individual chooses a random value s, computes $g(s)$ and sends to the pollster $\langle I_i, g(s)\rangle$.

- ***Decryption.*** Given an identifier I_i and a cipher-text c, the pollster decrypts c to recover the plaintext p, then computes the least significant bit b of $h(I_i \parallel p)$, and stores it. This bit b is a single bit of I_i's preference.

Since E and g have the same image, the pollster cannot discern which function was used to build the ciphertext c it receives. It can only decrypt it as described above, obtaining a plaintext, p. Only if c was computed using E do we have that $p = r$, and r contains only one bit of the individual's preference. Notice that the principal cannot decrypt $g(s)$ because she ignores the appropriate key.

However, although the pollster is not entitled to publish the preferences as such, it must be allowed to publish without any risk of indictment the statistical results obtained from them. To allow this, the function that was used to produce the statistical results must conform to the definition of *differential privacy* [4] (Definition 1).

Definition 1 *A randomized function f over data sets gives $\epsilon-$differential privacy if for any two data sets X_1 and X_2, which differ in at most one point, and $S \subseteq Range(f)$,*

$$\Pr[f(X_1) \in S] \leq \exp(\epsilon) \times \Pr[f(X_2) \in S]$$

The intuition behind this definition is simple. We may think of X_1 and X_2 as two databases that differ in only one record. A function f satisfies the definition of differential privacy if there is a similar probability (for small values of the ϵ parameter) that the respective applications of f to the two databases yield the same "feature" S.

Another desirable property for the pollster is that the statistic eliminates the baits so that, by publishing the results of the statistic, the pollster will expose no piece of evidence that the principals can exploit to indict it.

As an extreme, if the pollster used the constant function $f(x) = 1$, it would be on the safe side because the constant reveals no baits. But such a function would be useless for statistical purposes. Conversely, if the function also satisfies differential privacy, then its output would also be statistically significant.

In our application, we may think of X_1 as the set of clean individual preferences and of X_2 as the set of preferences enriched with the baits. So, the actual differences to consider are the baits. According to Definition 1, their being interleaved to the preferences will not significantly change the results of the statistics if the statistical function preserves differential privacy.

The literature confirms that most statistical functions, such as histogram function, principal components analysis and k-means clustering, can be constructed to be differential-privacy preserving [5]. However, these issues are outside the focus of this paper.

A dishonest pollster may publish the raw collected data, thus breaching the privacy of the individual who submitted them. If this is the case, that individual can start the indictment phase thanks to the clues that the baits provide. To succeed, the principal must show a number of valid exhibits of the form:

$$\langle I_i,\ s_i,\ b_i \rangle$$

where I_i is the individual's identity, s_i is the bait and b_i is the indicted bit. Let us denote by D the decryption function corresponding to E. An exhibit is valid *if and only if* the bit decrypted by the pollster, i.e. the least significant bit of $h(I_i \parallel D(g(s_i)))$ is equal to b_i. Notice that the notation for the encryption algorithm E and for the decryption algorithm D is simplified by omitting the cryptographic key. It is understood that anyone can apply E as it requires the pollster's public key, whereas only the pollster can apply D as it requires its own private key.

Two parameters are crucial to regulate the indictment phase and therefore should be pre-agreed out of band between the pollster and the participating community. One, indicated as n_0, is the validity threshold for the accusation. The individuals should advance a number of valid exhibits higher than n_0. It is interesting that different individuals can contribute to reaching that number, precisely those whose data the pollster putatively published. Since each exhibit is based on the value of a single bit, an individual might just be successful at guessing a valid exhibit. But n_0 reduces the probability that the individuals guess a sufficient number of valid exhibits to $\frac{1}{2^{n_0}}$.

Another important parameter is indicated as w_n and represents a sort of verdict's tolerance. The pollster can successfully contest the indictment by demonstrating that at least $(\frac{1}{2} - w_n)n$ of the alleged exhibits are invalid. Therefore, the pollster will need to invalidate as fewer than half the exhibits as defined by w_n. It proves that an exhibit is invalid by outputting $r_i = D(g(s_i))$, with a proof of correct decryption, and demonstrating that the least significant bit of $h(I_i \parallel r_i)$ is not b_i.

The ϵ parameter in Definition 1 is linked to n_0 and w_n. Precisely, "safe values of ϵ in turn depend on the values of n_0 and w_n that govern the indictment rules. These values must be chosen to permit a sufficient level of safe disclosure" [7, §5].

3 Analysing the Protocol under Realistic Threats

Our main aim is to analyse how the SEP protocol withstands a real threat model where each principal behave maliciously to achieve the maximum personal benefit from participating in the protocol.

Let us consider the following realistic scenario as an example. A pollster P claims a statistical investigation about the people who are interested in a life insurance contract. The individuals submit their preferences bundled with baits. Then, P collects the preferences, applies its chosen function (i.e. histogram, principal components analysis, etc.) that conforms to Definition 1. Finally, it publishes the output, that is the statistical investigation's results.

All principals have hitherto kept an honest behaviour, but after the publication of the results, some individuals may decide to accuse P for an unfounded privacy breach, aiming in fact at a refund for the putative violation. Also the opposite scenario is realistic, as it sees the pollster purposely breach the individuals' privacy by selling the clean collected preferences to an insurance company. What happens, if the pollster is unreachable for the accusation phase that would arguably follow?

Let us take an abstract analysis standpoint, and consider the possibility that a principal may successfully act maliciously as described above to reach his illegal aims. It means that she has some advantage against other principals. But SEP should not allow this because it aims at ensuring

fairness between the individuals and the pollster. Therefore, to verify whether SEP keeps its promised fairness, we analyse it in a realistic threat model, the *BUG* model [2]. BUG partitions the protocol participants into three sets according to whether they follow the protocol or not. For simplicity, we only outline the basics here, whereas a complete treatment can be found in [2].

- *Bad*: the user does not follow the protocol steps, and she works to obtain the best advantages from the communication. For example she cheats, frauds, impersonates others principals, and so on;
- *Ugly*: the user changes her behaviour to Bad or Good principal according to the context. For example, if the user sees that she can achieve any advantage, then she will change in Bad principal, otherwise she will keep a Good behaviour;
- *Good*: the principal always follows the protocol steps.

This level of detail is sufficient to our analysis outlined below. It is clear that each static picture of the network, depicting the principals with the messages they have sent or received up to that stage, can be characterized in terms of behaviour. For example, Lowe's attack sees the man in the middle acting as Bad, the end point B as Good and the initiator as Ugly. However, principals may change behaviour, so that other pictures may show a different partition.

We conduct our analysis showing how SEP counters scenarios with the various combinations of principals, especially those including bad ones.

We begin by analysing the behaviour of the submitting individuals. If they are good principals, then they submit their preferences correctly bundled with the baits, and attempt no dispute. If they are bad, they will attempt to build and publish a fake collection of preferences with baits, but the pollster will get by thanks to the n_0 and w_n parameters discussed above. What if the individuals are ugly? They may simply decide to take a good or bad behaviour according to their personal cost/benefit analysis, but as explained above, the pollster cannot be indicted if it kept the protocol.

This evaluation confirms that SEP is robust against the individuals' malicious behaviour. This is a useful feature because the present real world sees the individuals who may change their behaviour to achieve the maximum personal benefit leaving ethics behind. The present technological setting lets them easily and cheaply acquire hardware and skills to act maliciously.

It is perhaps more interesting to conduct the same analysis about the pollster's behaviour. If the pollster is good, it does not commit any privacy breach because it publishes the output of the function that is conform to Definition 1. Therefore, the individuals cannot indict it, as showed above. Whereas, if the pollster is bad, it collects the individuals' preferences and then it publishes or sells them to a third party. In this case, the individuals start the indictment phase but, as the pollster is dishonest, it does not participate to the indictment, for example it pretends a Denial of Service (DoS) attack [9, 11]. Therefore, it does not provide any proof of correct decryption of the baits. Hence, the individuals cannot achieve the exhibits to indict the pollster.

Finally, if the pollster is ugly, then it publishes correctly the statistical investigation's results, but at the same time it publishes or sells the individuals' preferences to a third party. Therefore, to avoid the indictment it keeps the same behaviour of a bad principal, or for example, it can start the indictment phase decrypting some baits, but as it is ugly, it may want to only decrypt an insufficient number of baits (that is, lower than n_0) and then pretend a DoS attack. It is clear that by acting so, the pollster will prevent its own indictment.

This evaluation shows that if it is the pollster to keep a malicious behaviour, the individuals are unable to indict it, because they do not have any sufficient exhibits for a judge. Our analysis somewhat surprisingly emphasizes that the pollster must collaborate to its own indictment by decrypting — with a proof of correct decryption — the baits. Would a dishonest pollster help its own accusation in practice? This form of collaboration may be unrealistic in many real-world scenarios. For example, the collaboration that SEP requires resembles a detective short of clues on a criminal scene who asks a suspected killer to exhibit the crime gun. This simile is accurate: the ability to decrypt the baits signifies having used the crime gun. Another example comes from the mentioned statistical investigation on life assurance. In that case, the pollster knows that the individuals do not have the exhibits, and hence can sell the data without participating in the subsequent indictment phase. No judge will have sufficient evidence to sentence it guilty.

The protocols should be strengthened so as to facilitate valid accusations even when the pollster fails to collaborate. In fact, our conclusion is that SEP as it stands violates the fairness requirement between its peers by giving a little advantage to the pollster. We have noticed that if the pollster is bad, it will realistically not want to help in indicting itself. The next section describes how to strengthen SEP in this direction.

4 Strengthening the Self Enforcing Privacy Protocol

Our evaluation indicated that the protocol provides some advantage to the pollster over the individuals. Since the pollster has the individuals' preferences and itself is an essential principal in its own indictment, it can practically avoid being indicted. This is the main motivation for the need to improve SEP. We would like to give the individuals

sufficient evidence that their data were sent to the pollster. This would provide the required fairness.

Our idea is to adopt digital signatures to provide such evidence. Of course, implementing such an idea requires a Public Key Infrastructure (PKI) [10], so that the pollster is equipped with a signature key pair and relative certificates, the signature creation algorithm S and the signature verification algorithm V[3].

Therefore, the individuals using V can verify the validity and the integrity of the signed data, and through the PKI they can verify the validity of the pollster's certificate, that is the validity of its signature key. It is clear that because a Global PKI is not available at present, requiring a PKI may limit the scope of the modifications proposed below. However, secure, global polls do not seem currently an issue.

Moreover, suppose each individual I capable of computing asymmetric cryptography by means of encryption algorithm E_I and decryption algorithm D_I. Our first attempt to strengthen SEP extends the two sending phases with an extra message. Precisely, when the individual sends private data or baits to the pollster, she waits for an acknowledgement (ack) message from it (Figure 2).

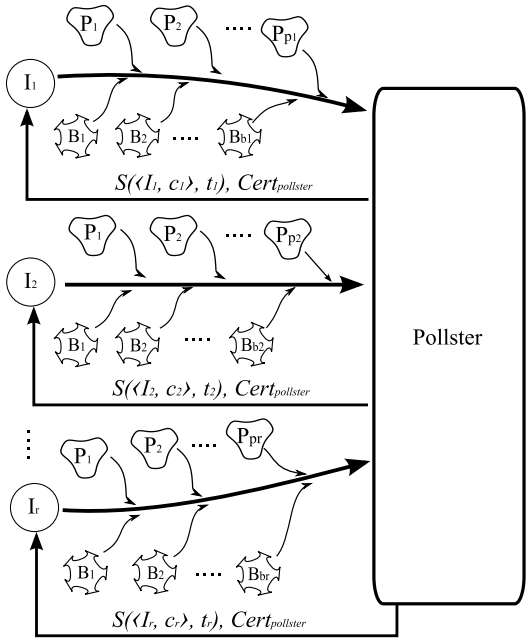

Figure 2. SEP+: our fairer variant of SEP

The form of the ack message is crucial. It shall be signed by the pollster to give it integrity and authenticity, and shall contain the outcome of the pollster's decryption. Therefore, the ack message is:

$$S(\langle I, c\rangle, t, E_I(D(c))), Cert_{pollster}$$

where I is the individual, t is the current timestamp, $Cert_{pollster}$ is the pollster's signature-key certificate, and c is:

- $E(r)$: if the individual sent a bit of her preference;
- $g(s)$: if she sent a bait.

Therefore, the novel ack message delivers the submitting individual the decryption that she may subsequently need for the indictment, that is $E_I(D(c))$.

When the individual receives the ack message, she verifies the certificate validity contacting a certification authority (CA), which belongs to the PKI. Then, with the right public key available, she verifies the digital signature using the V algorithm. If the pollster does not send the ack message, or the ack's sign is not correct, or the time-stamp is too old, then the individual will stop the communication. Incidentally, notice that for the certificate to be valid, the pollster must be registered with an accredited CA. This provides a reliable identification mechanism against the pollster also during an indictment phase that may follow.

It is worth remarking that the ack messages provide an individual with the decryption of the data that she sent. Therefore, if she sent a bit of preference, then she will receive r; otherwise if she sent a bait, then she will receive the decryption of $g(s)$, which she did not know otherwise. The decrypted value can be read only by that individual, because it is encrypted using E_I.

We evaluate also our updated protocol in the BUG threat model, especially to assess whether the principal can have some advantage over each other. If the pollster P is bad and it publishes or sells the collected data to a third party, then the individuals start the indictment phase. If P does not participate here, then they will be able to indict it all the same, because they have the decrypted baits collected from the ack messages. Therefore, they have a sufficient number of valid exhibits. Conversely, if P is good and the individuals are bad, they may make a fake collection using the decrypted values of the baits, publish it and then accuse P. In this case, P will be unfairly indicted because the individuals knew the baits in advance.

Therefore, this updated protocol removes the advantage from the pollster but moves it to the individuals. In consequence, our update fails to make the original protocol fair. It is now clear that it is inappropriate to give the decrypted baits to the individuals before any actual indictment. To achieve more fairness, we advance a different update to the protocol, resulting in what we address as SEP+.

Also SEP+ assumes a PKI with signature creation algorithm S and signature verification algorithm V for the pollster, but none are needed for the individuals. But SEP+ extends the original sending phases with a simple ack message (simpler than the previous attempt) of the form:

$$S(\langle I,\ c\rangle, t), Cert_{pollster}$$

It can be observed that this message does not provide the decrypted c, that is the pollster simply replies by signing the just received pair along with the current timestamp. As with the failed variant, the individuals will continue the protocol only if the pollster sends the ack correctly and timely, otherwise they will abort the session.

SEP+ must be realistically evaluated against the BUG threat model. As with the original protocol, if the individuals are bad and the pollster P is good, then they will make a fake collection, publish it and finally attempt to accuse P. Because they do not know the decrypted values of the baits, they are unable to indict P illicitly, exactly as with SEP.

The new portion of the evaluation is the converse scenario. Let us consider a bad pollster P running SEP+ with the submitting individuals. The novelty is that, as P publishes the collected preferences, the individuals can still indict P even if it does not participate actively in its indictment. They have collected the ack messages that are signed by P. This means that the individuals have evidence that their data where received by P. In particular, P's signature may have been pre-agreed to signify that the pollster accepts compliance with the individuals' privacy preservation policy. Thus, the ack messages qualify as valid clues that the individuals can show to the judge if the pollster committed a breach and then did not want to participate in its own indictment.

SEP+ inherits the indictment phase from the SEP protocol, and therefore it would be ideal if the pollster contributed actively in the indictment phase. However, even if the pollster does not participate, its signed ack messages qualify as sufficiently probatory evidence thanks to the robustness of digital signature. Therefore, SEP+ narrows down the pollster's malicious behaviour balancing fairness towards the individuals.

5 Conclusions

The scheme presented by Golle et al. [7] intuitively provides a simple method to resolve the problem of privacy preservation in e-polls scenarios. It provides some level of fairness between its participants, even and especially when they misbehave. In particular, we selected only one of their techniques ([7, §5]), which we addressed as SEP, as it seems to achieve the highest level of fairness against a comparatively acceptable simplicity.

Our paper analysed SEP (§3) in a real threat model. We showed that the protocol does not achieve sufficient fairness against the realistic threat of a malicious pollster that arguably will not help out during the indictment phase against itself. Strictly speaking, SEP puts the pollster in a somewhat advantageous situation over the individuals.

In this light, we advanced two variants for the protocol. The first variant was found to move the advantages from the pollster over the individuals, and thus did not reach a satisfactory level of fairness. The second variant, addressed as SEP+, reached that level by giving the submitters a digital signature that the pollster received their preferences. The underlying PKI helps pinpoint the pollster's operations because the pollster is registered with some certification authority. Our future work includes simulating SEP+ as for data and transmission overhead using a mechanised network simulator such as NS2 [1].

The most general outcome of our research is that also recently published privacy protocols such as SEP must be analysed in a currently realistic threat model. It is the same lesson that was learnt empirically during the early 1990s with classical security protocols such as Needham-Schroeder and Woo-Lam. Our research confirms that Computer Security is sufficiently mature to easily and quickly recast that lesson in the context of privacy.

Acknowledments This work was supported by the FP7-ICT-2007-1 Project no. 216471, "AVANTSSAR: Automated Validation of Trust and Security of Service-oriented Architectures" (http://www.avantssar.eu).

References

[1] The network simulator - ns 2 - http://www.isi.edu/nsnam/ns/.

[2] G. Bella, S. Bistarelli, and F. Massacci. Retaliation: Can we live with flaws? In *Proc. of the Nato Advanced Research Workshop on Information Security Assurance and Security*, 2005.

[3] R. Cramer and V. Shoup. Signature schemes based on the strong RSA assumption. *ACM Transactions on Information and System Security*, 3:161 – 185, 2000.

[4] C. Dwork. Differential privacy. In *ICALP 2006*, 2006.

[5] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, 2006.

[6] E. Fujisaki, T. Okamoto, D. Pointcheval, and J. Stern. RSA–OAEP is secure under the RSA assumption. In *Advances in Cryptology — Proceedings of CRYPTO '2001*, 2001.

[7] P. Golle, F. McSherry, and I. Mironov. Data collection with self-enforcing privacy. In *Proceedings of the 13th ACM conference on Computer and communications security*, 2006.

[8] B. Kaliski and J. Staddon. RSA cryptography specifications version 2.0 - http://www.ietf.org/rfc/rfc2437.txt.

[9] F. Lau, S. H. Rubin, M. H. Smith, and L. Trajkovic. Distributed denial of service attacks. In *IEEE International Conference on Systems, Man, and Cybernetics*, 2000.

[10] U. Maurer. Modelling a public-key infrastructure. In *ESORICS: European Symposium on Research in Computer Security*, 1996.

[11] C. L. Schuba, I. V. Krsul, M. G. Kuhn, E. H. Spafford, A. Sundaram, and D. Zamboni. Analysis of a denial of service attack on TCP. In *Proceedings of the 1997 IEEE Symposium on Security and Privacy*, 1997.