Coded Modulation and the Arrival of Signcryption¹

Yuliang ZHENG

Department of Software and Information Systems, UNC Charlotte 9201 University City Blvd, Charlotte, NC 28223, USA Email: yzheng@uncc.edu

Abstract. Digital communications engineers strive to transmit data in a more efficient and reliable manner, while cryptographers endeavor to offer both data confidentiality and integrity. In the 1970's and 1980's, researchers in digital communications successfully developed techniques for multi-level coded modulation that simultaneously allow the correct of transmissions errors and the demodulation of signals upon the arrival at the receiver. Reminiscent of coded modulation, signeryption is a cryptographic technique developed to "hit two birds in one stone", namely to provide both data confidentiality and origin unforgeability with reduced overhead. In this article the author recounts his unique experience in witnessing the evolution and perfection of coded modulation techniques, and more important, how the experience inspired him to embark on the journey of searching for efficient techniques for combining public key encryption and digital signatures; the author also attempts to illuminate future directions for efficient, hybrid cryptographic techniques.

Keywords. Coded modulation, digital signature, hybrid encryption, public key cryptography, signcryption

1. Coded Modulation

In a typical communications system, data from an originator undergoes a sequence of transformations prior to being transported to its intended recipient. These transformations may include *source encoding* to compress the data or remove unwanted redundant information from the data, *authentication tagging* to ensure the detection of unauthorized modification, *encryption* to prevent the data from being accessible to unauthorized parties while en route, *error correction encoding* to allow the recipient to detect and correct transmission errors, and finally *modulation* of data signals for transmission over a communications channel between the originator and the recipient. Generally the communications in reserve order. Figure 1 depicts the various operations on data while traveling across a communications channel. Note that in the figure, authentication is applied before encryption is carried out. Alternatively, encryption can be applied first, followed by authentication.

¹Invited talk at NATO Advanced Research Workshop on Enhancing Crypto-Primitives with Techniques from Coding Theory, 6-9 October, 2008, Veliko Tarnovo, Bulgaria.

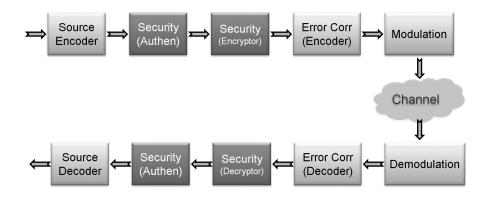


Figure 1. A Communications System

Error correcting codes and modulation techniques have been at the core of digital communications engineering and research from the mid 20th century. As a typical communications channel has only a limited bandwidth, one of the central questions is how to minimize the loss of effective data transmission rate incurred by error correcting codes. Another important question is how to reap the full benefits of an increased data transmission rate offered by multi-level/phase modulation without suffering worsening interference among signals. While a great number of error correcting codes and modulation techniques had been discovered accompanying the advent of digital communications after the Second World War, historically, error correction and modulation had always been carried out separately.

During the 1970's, researchers embarked on the pursuit of techniques for combining error correction and modulation with the aim of achieving performance gain without requiring the expansion of bandwidth or incurring a significant reduction of data transmission rate. The most successful of those efforts was represented by the work of Ungerboeck [1,2,3,4], and independently, of Imai and Hirakawa [5] (see also [6]). Ungerboeck focused on blending together trellis or convolution codes and multi-level modulation without sacrificing bandwidth efficiency, whereas Imai and Hirakawa had the same goal but used a different approach, which was to combine block error correcting codes and multi-level modulation.

Coded modulation addresses simultaneously two issues that appear mutually contradictory: (1) high transmission reliability, and (2) high transmission efficiency. The hybrid technique, whether it is based on trellis codes or block error correcting codes, makes reliable and bandwidth efficient data transmission a reality (see Figure 2).

2. Musings on Blending

The 1980's was an exciting period for those who worked in telecommunication. At the time, Professor Hideki Imai led a number of research projects at Yokohama National University in Japan. These projects covered virtually all important technical aspects pertaining to data processing and communications. Specifically, the research projects addressed source coding, cryptography, error correcting codes, modulation and coded modulation.

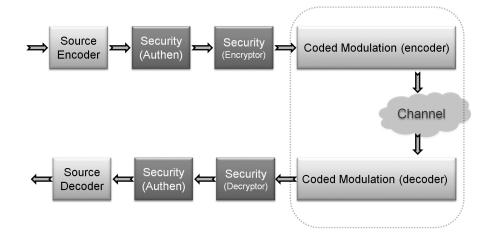


Figure 2. Coded Modulation for Bandwidth Efficient Communications

I joined Professor Imai's group to pursue my graduate studies in the mid 1980's. Although my research area was cryptography, I was fortunate to be able to participate in weekly seminars with fellow students who worked on a range of different research projects. I still vividly remember that at one of the long seminars that extended well into the late evening, a fellow student started to explain the Imai-Hirakawa coded multi-level modulation technique [5] when I felt somewhat tired after a long day's studies and discussions. I was immediately intrigued by the idea of blending error correcting codes and modulation together to obtain a better solution than applying them separately. Over the next few years I maintained strong interest in coded modulation. What I felt most fascinating was not only the beauty of ideas behind the technology but also the amazing velocity at which the technology was perfected, standardized and applied in practice.

As a cryptographic researcher witnessing the rapid maturing and adoption of coded modulation in digital communications, I asked myself a quite natural question, that was whether it was feasible to combine two cryptographic primitives into something that would be more efficient than employing the two primitives separately. The question accompanied me during the entire remaining period of my graduate studies. Years after I finished my PhD and moved Down Under, I found it hard for me to stay completely away from musings on the same question.

Two of the most important functions of modern cryptography are the assurance of data confidentiality and that of data integrity. Confidentiality can be achieved using encryption algorithms or ciphers, whereas integrity can be provided by the use of authentication techniques.

Encryption algorithms fall into one of two broad groups: private key encryption and public key encryption. Likewise, authentication techniques can be categorized by private key authentication algorithms and public key digital signatures. When examining a cryptographic algorithm, one needs to take into account not only the strength or level of security the algorithm can offer, but also the computational time it takes to perform the algorithm, together with the message expansion incurred by the algorithm. When two cryptographic algorithms offer a similar level of security, computational time and message expansion become a focal point of comparison. As a rule, smaller computational time and shorter message expansion are generally considered more desirable.

While both private key encryption and private key authentication admit very fast computation with minimal message expansion, public key encryption and digital signatures generally require heavy computation, such as exponentiations involving very large integers, together with message expansion proportional to security parameters (such as the size of a large composite integer or the size of a large finite field). Figure 3 illustrates the computational and message overhead incurred when digital signatures and public key encryption are applied in succession to a message.

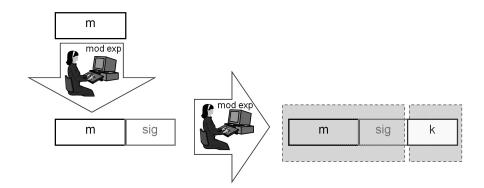


Figure 3. Digital Signature Followed by Public Key Encryption

I realized that there were at least two types of combinations that appeared to be meaningful in practice:

- 1. private key encryption combined with private key authentication, and
- 2. public key encryption combined with digital signatures.

I also realized that the most important goal of a successful combination of two different cryptographic algorithms should be for the resultant solution not only to be faster to compute but also to admit shorter message expansion, when compared to applying the two original algorithms separately.

I set my sight on the second type of combinations, namely combinations of public key encryption and digital signatures, for a number of reasons:

- 1. Future widespread use of battery powered small devices such as smart cards, smart phones, personal digital assistants (PDAs), electronic passports, electronic wallets and other types of gadgets would require new public key cryptographic techniques that consume as little battery power as possible.
- Applications in a resource constrained environment such as contactless wireless identification tokens and unattended remote data collection systems would require the use of public key cryptographic algorithms that are not only fast to compute but also introduce minimum data expansion.
- 3. Finding combined public key cryptographic solutions appeared more challenging.

4. I felt that my experience in designing techniques for "immunizing" public key encryption against chosen ciphertext attacks [7,8] would be useful in addressing the new challenge. The essence of these "immunization" techniques was to use authentication tags, especially those generated by a (keyed) one-way hash algorithm, to transform an unstructured plaintext into a highly structured one prior to the application of public key encryption. The transformation incapacitates a chosen ciphertext attacker who attempts to create a new ciphertext without already knowing the corresponding plaintext. Some of the ideas were later elaborated as the "random oracle" model and plaintext awareness by other researchers, and constituted an important foundation for public key encryption that admitted provable security.

3. Signcryption

After setting as my goal the combination of public key encryption and digital signatures, I narrowed down candidate algorithms to ElGamal public key encryption and signature, especially those that relied for their security on the hardness of discrete logarithm on a *subgroup* of a large finite field, due to their outstanding efficiency as well as the availability of their counterparts on elliptic curves.

3.1. ElGamal Public Key Encryption and Signature in a Subgroup

The specific version of ElGamal public key encryption and digital signatures I was interested in involved three parameters that were public to all:

- 1. *p*: a large prime.
- 2. q: a prime factor of p-1.
- 3. g: an integer in the range of $[1, \ldots, p-1]$ with order q modulo p.

Consider two users Alice and Bob. Alice has a private key x_a chosen uniformly at random from $[1, \ldots, q-1]$. She also has $y_a = g^{x_a} \mod p$ as her matching public key. Likewise, Bob's private key is an integer x_b chosen uniformly at random from $[1, \ldots, q-1]$, and his public key is $y_b = g^{x_b} \mod p$.

Now assume that Alice wishes to send a message m to Bob in a secure manner. Alice first looks up Bob's public key y_b in a public key directory. She then picks a random integer x from $[1, \ldots, q-1]$, and calculates $t = y_b^x \mod p$. This is followed by employing an appropriate one-way hash algorithm hash to compute from t an encryption key k = hash(t) for an appropriate private key cipher (E, D). Finally Alice sends to Bob the following pair of data items as a ciphertext of m:

ElGamal Encryption : $(c_1, c_2) = (g^x \mod p, E_k(m))$

where E is the encryption algorithm of the private key cipher.

Upon receiving (c_1, c_2) , Bob can recover k by computing $k = hash(c_1^{x_b} \mod p)$. He can then use k and the decryption algorithm D of the same private key cipher to decrypt c_2 and obtain m.

Alice's signature on a message m is composed of two numbers r and s which are defined as

ElGamal Signature : $(r, s) = (g^x \mod p, (hash(m) - x_a \cdot r)/x \mod (p-1))$

where x is a random number picked from [1, ..., q - 1], and *hash* is an appropriate one-way hash algorithm. Bob and any other party can verify the authenticity of Alice's signature on the message m by using her publicly available key y_a .

There are numerous variants of and improvements to the original ElGamal signature. The most notable ones include the NIST digital signature standard (DSS) or digital signature algorithm (DSA) [9], and the Schnorr signature [10]. These two signature techniques are defined as

$$DSS: (r, s) = ((g^x \mod p) \mod q, (hash(m) + x_a \cdot r)/x \mod q)$$

Schnorr: $(r, s) = (hash(g^x \mod p, m), (x - x_a \cdot r) \mod q)$

Two more interesting variants are obtained by further shortening variants of the DSS. These two shortened versions are called SDSS1 and SDSS2 respectively, and are defined as:

$$SDSS1: (r, s) = (hash(g^x \mod p, m), x/(r + x_a) \mod q)$$
$$SDSS2: (r, s) = (hash(g^x \mod p, m), x/(1 + x_a \cdot r) \mod q)$$

3.2. Basic Signcryption Algorithms

Looking closely at the ElGamal encryption and signature algorithms, one notices that both contains the following item:

$$g^x \mod p$$

This quantity can be viewed as playing the role of an "ephemeral key" in both algorithms. An interesting question is *whether it is possible to let the same "ephemeral key" serve as a conduit linking the encryption and signature algorithms together.*

Further, one notices that $g^x \mod p$ does not explicitly appear in any of the four variants of the ElGamal signature described above. Nevertheless the quantity can be easily derived from these signatures by a signature verifier. All these four variants exhibit a significantly shorter signature size than the original ElGamal signature. This brings up yet another interesting question, that is whether it is possible to combine ElGamal encryption and signature in such a way that the resultant algorithm does not contain $g^x \mod p$.

After a number of trials and errors, I was able to firm up my thinking in the southern winter of 1996. The outcome, which I called "signeryption", was a nice combination of ElGamal encryption and signature that answered both questions above in the affirmative. I will explain the combination that is based on SDSS1. In describing the signeryption technique, I use Hash to denote a one-way hash algorithm, KH a keyed one-way hash algorithm, and (E, D) a private key cipher.

Basic Algorithm Signcryption of m by Alice the Sender

- Pick x uniformly at random from [1,...,q-1], and let k = hash(y^x_b mod p). Split k into k₁ and k₂ of appropriate length.
 c = E_{k1}(m).
 r = KH_{k2}(m).
 s = x/(r + x_a) mod q.
- 5. Output (c, r, s) as the ciphertext to be sent to Bob.

Basic Algorithm Unsigncryption of (c, r, s) by Bob the Recipient

- 1. Recover k from r, s, g, p, y_a and x_b
- by $k = hash((y_a \cdot g^r)^{s \cdot x_b} \mod p)$. 2. Split k into k_1 and k_2 .
- 3. $m = D_{k_1}(c)$.
- 4. Output m as a valid message originated from Alice
- only if $KH_{k_2}(m) = r$. Output "Reject" otherwise.

The technique was first detailed in a patent application in Oct. 1996 [11], although the research paper was not published almost a year later at Crypto'97 [12].

3.3. Provably Secure Signcryption

Following the publication of the basic signcryption algorithm discussed above, finding formal proofs for both confidentiality and unforgeability of the algorithm emerged as the next challenge. I was fortunate to have Ron Steinfeld joining my lab at Monash University as a PhD student in the southern fall of 1999. Ron took my advice to look into formal proofs for the security of signcryption. We soon realized that identifying a right security model for signcryption was of most importance. In late 1999 we made the first step in that direction: we were able to find a security proof for the unforgeability of a factoring based signcryption algorithm. The result was presented at ISW2000 [13], leaving proofs for the confidentiality of the signcryption algorithm as an open problem.

I welcomed Joonsang Baek to join my lab as a PhD student in early 2000. Shortly after his arrival, Joonsang started to work with Ron and myself on security proofs for signcryption. The joint research turned out to be extremely fruitful, resulting in the establishment of a strong security model for signcryption in the multi-user setting and formal security proofs for both unforgeability and confidentiality in that model [14,15].

Independent of work at my lab, An, Dodis and Rabin succeeded in obtaining proofs for the security of a broad class of joint public key encryption and digital signatures in the two-user setting [16]. These results and our results for the multi-user setting were mutually complementary, representing an important step towards designing signcryption that admits provable security. The original signcryption algorithm required a few tweaks in order for its security to be proved with mathematical rigor [14,15]. The tweaked algorithm employs two separate one-way hash algorithms G and H. The former was used to generate a key for a private key cipher whereas the latter to compute the value of r. In addition, both Alice's public key and Bob's public key participated in the hash computation of r, whereby the ciphertext was tightly bound to both Alice and Bob, thwarting possible abuse by dishonest Alice or Bob.

Provably Secure Algorithm Signcryption of m by Alice the Sender

- 1. Pick x uniformly at random from $[1, \ldots, q-1]$.
- 2. $k = y_b^x \mod p$.
- 3. $\tau = G(k)$.
- 4. $c = E_{\tau}(m)$.
- 5. $r = H(m, y_a, y_b, k)$.
- 6. If $r + x_a = 0 \pmod{q}$ then go back to Step 1;
- otherwise let $s = x/(r + x_a) \mod q$.
- 7. Output (c, r, s) as the ciphertext to be sent to Bob.

Provably Secure Algorithm Unsigncryption of (c, r, s) by Bob the Recipient

- 1. $k = (y_a g^r)^{s \cdot x_b} \mod p$.
- 2. $\tau = G(k)$.
- 3. $m = D_{\tau}(c)$.
- 4. If $H(m, y_a, y_b, k) = r$ then output m as a valid message originated from Alice; otherwise output "Reject".

4. Extensions, Standardization and Future Research Directions

Signcryption has since been extended to elliptic curves [17], integer factorization [13, 18,19], and pairings [20]. Furthermore researchers have designed numerous signcryption techniques that have additional useful properties such as insider/outsider security [21, 22], direct verifiability by a third party [23], threshold [24], blindness [25], identity as a public key [26], certificateless [27], proxy [28], and many others. The reader is directed to "Signcryption Central" (www.signcryption.net) which serves as an information portal for recent developments in the field [29].

In a different direction, Jutla studied the integration of private key encryption and private key message authentication, giving rise to *authenticated encryption* or *authencryption* [30].

More recently, significance of signcryption in real world applications has gained recognition by experts in data security. Since 2007, a technical committee within the International Organization for Standardization (ISO/IEC JTC 1/SC 27) has been devel-

oping an international standard for signcryption techniques [31]. Techniques to be included in the standard must meet ISO's stringent requirements, especially those pertinent to security, performance and maturity.

To close the article, I would like to bring the reader's attention to Figure 4 which depicts a communications system where both coded modulation and signcryption are employed, achieving gains not only in communications efficiency and reliability, but also in data security, all with minimal overhead. One cannot help but ask: are additional types of blending still possible?

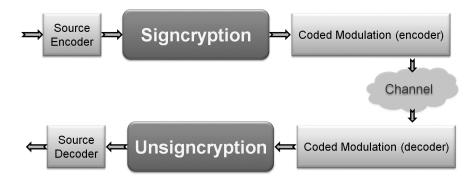


Figure 4. Signcryption in a Communications System

Acknowledgements

I dedicate this article to Shuji Hirakawa, Hideki Imai and Gottfried Ungerboeck, whose pioneering work in coded modulation inspired me to develop signcryption. I am also grateful to Alex Dent for his insightful comments and suggestions on the article.

References

- Gottfried Ungerboeck and I. Csajka. On improving data-link performance by increasing the channel alphabet and introducing sequence coding. In *Proceedings of 1976 International Symposium on Information Theory*, Ronneby, Sweden, June 1976.
- [2] Gottfried Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, 28(1):56–67, 1982.
- [3] Gottfried Ungerboeck. Trellis-coded modulation with redundant signal sets, part I: Introduction. *IEEE Communications Magazine*, 25(2):5–11, 1987.
- [4] Gottfried Ungerboeck. Trellis-coded modulation with redundant signal sets, part II: State of the art. IEEE Communications Magazine, 25(2):12–21, 1987.
- [5] Hideki Imai and Shuji Hirakawa. A new multilevel coding method using error-correcting codes. *IEEE Transactions on Information Theory*, 23(3):371–377, 1977.
- [6] Kazuhiko Yamaguchi and Hideki Imai. A study on Imai-Hirakawa trellis-coded modulation schemes. In AAECC-6: Proceedings of the 6th International Conference, on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, pages 443–453, London, UK, 1989. Springer-Verlag.
- [7] Yuliang Zheng and Jennifer Seberry. Spractical approaches to attaining security against adaptively chosen ciphertext attacks. In Advances in Cryptology - CRYPTO'92, volume 740 of Lecture Notes in Computer Science, pages 292–304, Berlin, New York, Tokyo, 1993. Springer-Verlag.

- [8] Yuliang Zheng and Jennifer Seberry. Immunizing public key cryptosystems against chosen ciphertext attacks. *IEEE Journal on Selected Areas in Communications*, 11(5):715–724, June 1993.
- [9] National Institute of Standards and Technology. Digital signature standard (DSS). Federal Information Processing Standards Publication FIPS PUB 186, May 1994; FIPS PUB 186-2, Jan. 2000; FIPS PUB 186-3, Nov. 2008, U.S. Department of Commerce.
- [10] Claus-Peter Schnorr. Efficient identification and signatures for smart cards. In Advances in Cryptology - CRYPTO'89, volume 435 of Lecture Notes in Computer Science, pages 239–251, Berlin, New York, Tokyo, 1990. Springer-Verlag.
- [11] Yuliang Zheng. Message encryption and authentication methods (signcryption). Australia Patent Serial Number 721497, lodged on October 25, 1996, granted on May 10, 2000; US Patent 6,396,928, granted on May 28, 2002.
- [12] Yuliang Zheng. Digital signcryption or how to achieve cost(signature & encryption) << cost(signature) + cost(encryption). In Advances in Cryptology - CRYPTO'97, volume 1294 of Lecture Notes in Computer Science, pages 165–179, Berlin, New York, Tokyo, 1997. Springer-Verlag.
- [13] Ron Steinfeld and Yuliang Zheng. A signcryption scheme based on integer factorization. In Information Security — Proceedings of 2000 Information Security Workshop (ISW'00), volume 1975 of Lecture Notes in Computer Science, pages 308–322, Berlin, New York, Tokyo, 2000. Springer-Verlag.
- [14] Joonsang Baek, Ron Steinfeld, and Yuliang Zheng. Formal proofs for the security of signcryption. In Proceedings of 2002 International Workshop on Practice and Theory in Public Key Cryptography (PKC2002), pages 80–98, London, UK, 2002. Springer-Verlag.
- [15] Joonsang Baek, Ron Steinfeld, and Yuliang Zheng. Formal proofs for the security of signcryption. *Journal of Cryptology*, 20(2):203–235, 2007.
- [16] Jee Hea An, Yevgeniy Dodis, and Tal Rabin. On the security of joint signature and encryption. In Advances in Cryptology - EUROCRYPT'02, volume 2332 of Lecture Notes in Computer Science, pages 83–107, Berlin, New York, Tokyo, 2002. Springer-Verlag.
- [17] Yuliang Zheng and Hideki Imai. Efficient signcryption schemes on elliptic curves. In *Global IT Security* — *Proceedings of the IFIP TC11 14th International Conference on Information Security (IFIP/SEC'98)*, pages 75–84, Vienna and Budapest, August 1998. International Federation for Information Processing (IFIP).
- [18] Yuliang Zheng. Identification, signature and signcryption using high order residues modulo an RSA composite. In *Proceedings of 2001 International Workshop on Practice and Theory in Public Key Cryptography (PKC2001)*, volume 1992 of *Lecture Notes in Computer Science*, pages 48–63, Berlin, New York, Tokyo, 2001. Springer-Verlag.
- [19] John Malone-Lee and Wenbao Mao. Two birds one stone: Signcryption using RSA. In *Topics in Cryptology CT-RSA 2003*, volume 2612 of *Lecture Notes in Computer Science*, pages 211–225, Berlin, New York, Tokyo, 2003. Springer-Verlag.
- [20] Benoit Libert and Jean-Jacques Quisquater. New identity based signcryption schemes from pairings. In Proceedings of 2003 IEEE Information Theory Workshop, pages 155–158, 2003.
- [21] Alexander W. Dent. Hybrid signcryption schemes with insider security. In Proceedings of 10th Australasian Conference on Information Security and Privacy, volume 3574 of Lecture Notes in Computer Science, pages 253–266, Berlin, New York, Tokyo, 2005. Springer-Verlag.
- [22] Alexander W. Dent. Hybrid signcryption schemes with outsider security. In Proceedings of the 8th International Information Security Conference (ISC'05), volume 3650 of Lecture Notes in Computer Science, pages 203–217, Berlin, New York, Tokyo, 2005. Springer-Verlag.
- [23] Feng Bao and Robert H. Deng. A signcryption scheme with signature directly verifiable by public key. In PKC '98: Proceedings of the First International Workshop on Practice and Theory in Public Key Cryptography, volume 1431 of Lecture Notes in Computer Science, pages 55–59, London, UK, 1998. Springer-Verlag.
- [24] Shanshan Duan, Zhenfu Cao, and Rongxing Lu. Robust id-based threshold signcryption scheme from pairings. In Proceedings of the 3rd international conference on Information security (InfoSecu'04), volume 85 of ACM International Conference Proceeding Series, pages 33–37. ACM, 2004.
- [25] Tsz Hon Yuen and Victor K. Wei. Fast and proven secure blind identity-based signcryption from pairings. In *Topics in Cryptology — CT-RSA 2005*, volume 3376 of *Lecture Notes in Computer Science*, pages 305–322, Berlin, New York, Tokyo, 2005. Springer-Verlag.
- [26] Xavier Boyen. Multipurpose identity-based signcryption: A swiss army knife for identity-based cryptography. In Advances in Cryptology - CRYPTO'03, volume 2729 of Lecture Notes in Computer Science,

pages 383-399, Berlin, New York, Tokyo, 2003. Springer-Verlag.

- [27] Manuel Barbosa and Pooya Farshim. Certificateless signcryption. In Proceedings of the 2008 ASIAN ACM symposium on Information, computer and communications security (ASIA-CCS'08), pages 369– 372. ACM, 2008.
- [28] Chandana Gamage, Jussipekka Leiw, and Yuliang Zheng. An efficient scheme for secure message transmission using proxy-signcryption. In *Computer Science — Proceedings of the 22nd Australasian Computer Science Conference (ACSC'99)*, volume 21 of *Australian Computer Science Communications*, pages 420–431. Springer-Verlag, 1999.
- [29] Signcryption Central. www.signcryption.net.
- [30] Charanjit S. Jutla. Encryption modes with almost free message integrity. In Advances in Cryptology - EUROCRYPT'01, volume 2045 of Lecture Notes in Computer Science, pages 529–544, Berlin, New York, Tokyo, 2001. Springer-Verlag.
- [31] ISO/IEC 29150. Information technology Security techniques Signcryption. 2008. (Under development).