Use of ID-based Cryptography for the Efficient Verification of the Integrity and Authenticity of Web Resources

Thanassis Tiropanis[1] and Tassos Dimitriou[2]

[1] University of Southampton, UK,
tt2@ecs.soton.ac.uk
[2] Athens Information Technology, Greece,
tdim@ait.edu.gr

Abstract. As the amount of information resources on the Web keeps increasing so are the concerns for information integrity, confidentiality and authenticity. In Web 2.0 users are producers as well as consumers of content and metadata, which makes guaranteeing the authenticity and integrity of information critical. The scale of the Web requires that any proposals in this direction require minimal (if any) infrastructural or administrative changes. This paper proposes the use of ID-based cryptography (IBC) to address requirements for integrity and authenticity of Web resources using either the URL/URI of a resource or the DNS name part of if. This approach presents certain challenges, which are discussed along with the pros and cons of different designs and implementations.

Key words: Identity Based Cryptography, Integrity, Authenticity, Web 2.0

1 Introduction

The number of Internet and Web users has been increasing at very high rates over the last decade and, considering that Internet penetration in developing countries is still relatively low there is space for ongoing increase in the coming years. The amount of content that is exchanged is constantly growing and a number of content distribution and delivery infrastructures (peer-to-peer, content repositories) are proving particularly popular. Web 2.0 applications allow users to be not just content consumers but also producers. The importance of content in this emerging paradigm of Web-service deployment and use has already been identified; according to O'Reilly [10] "data is the next Intel inside".

Currently there is research in progress on a new content-centric communication paradigm that aspires to transform networking by focusing not on enabling the communication between network end-points but on identifying content to be obtained from networks using client-server, peer-to-peer or other types of exchanges. There are expectations that this effort will lead to more efficient networks in terms of content distribution and more efficient services ([9], [11]).

This vision of content-centric communication however, is based on the assumption that a network infrastructure will be able to (i) identify each Web resource uniquely and (ii) provide guarantees on the integrity and authenticity of Web resources since they can be obtained not exclusively from their source network end-point but over a peer-to-peer or other type of content distribution network. The requirements of integrity and authenticity are increasingly critical as the content currently produced by an already large user base is set to keep growing. To this end we propose the use of Identity-Based Cryptography (IBC) as an efficient and scalable way of guaranteeing the authenticity and integrity of Web resources using an IBC-based system for efficiently signing and verifying Web based content. Our proposal involves the use of the URL/URIs (or the DNS name part of URL/URIs) of resources as IBC identifiers of every resource to be disseminated over the Web, Peer-to-Peer or other content dissemination networks combined. We also propose the use of identity-based digital signatures.

The proposed approach does not require infrastructural changes and we believe it can therefore be seamlessly introduced, making use of the existing XML Signature Syntax and Processing Standard of the W3C [14].

The rest of the paper is organized as follows: Section 2 reviews the existing literature on Identify-Based Cryptography (IBC), on Web 2.0 applications and on content dissemination infrastructure. Section 3 describes in detail how IBC can provide an efficient and scalable way to address integrity and authenticity concerns in content-centric communication. Section 4 discusses different deployment scenarios over the existing Web infrastructure, while Section 5 provides a discussion on the proposed approach and identifies further work items and related research directions.

2 Background

2.1 Supporting Identity Based Cryptography (IBC) on the Internet

Identity-based cryptography (IBC) was first introduced by Shamir back in 1984 [13]. While the original scheme of Shamir supported only signature operations, recently, there has been an increased interest in the use of IBC which was due to the discovery of a secure Identity Based Encryption scheme based on pairings over elliptic curves by Boneh and Franklin [1].

In an identity-based cryptosystem, public keys can be derived from arbitrary strings while the corresponding private keys are generated and distributed by an associated Trusted Authority (TA). Thus an identity-based cryptosystem enjoys most of the benefits of public key cryptography without the need for certificates and the problems they present. This in turn leads to a more lightweight approach to deploying public key cryptography [12]. In the sequel we review some basic IBC systems that have been proposed to date.

Boneh *et al.* [2] proposed a new approach to certificate revocation centered around the concept of an on-line SEmi-trusted Mediator called the SEM. The use of the SEM in conjunction with a simple threshold variant of the RSA

cryptosystem enables the quick revocation of all security capabilities of a user. The proposal of mediated RSA was then used in an identity based setting in [3] as a simple solution and alternative to the Weil pairing scheme of [1].

One of the advantages of PKIs is that they can be organized into hierarchies which reflect the internal structure of a large organization or group. Recent work, however, has demonstrated the ability to implement similar hierarchies in an IBC context ([6, 7]). In [17], this is taken one step further and an attempt was made to integrate this approach into existing standards and software, so as to ease deployment.

Finally, Crampton *et al.* [4] discuss how various Identity-based cryptographic techniques can be used to provide web services security. In particular, the authors compare Identity-based with traditional, certificate based techniques and they show how the first type can be used to secure XML messages in a more lightweight way compared to the second one.

2.2 Digital signatures for Web resources

The need to provide digital signatures for Web resources has been identified by the W3C and the IETF which engaged in common standardization activity (www.w3.org/Signature) and compiled requirements for XML Signatures in terms of a data model, syntax, format and processing (RFC2807 [15]). The core standard to emerge from this activity is the "XML Signature Syntax and Processing" [14] standard, which provides for digital signatures as XML documents.

XMLDsig can be used to sign not only XML documents but also resources in other formats. To verify the authenticity and integrity of a resource, one needs to have or obtain the key to be used; in a PKI setting, this is the public key of the entity that signed the resource. Considering that a large number of Web resources are specified in XML and that the use of the XML-compatible version of HTML (xHTML) is widely used today, it seems that XMLDsig provides a number of ways to package signatures into a large number of Web resources without changes to existing Web infrastructure.

3 IBC for Web resources

Our proposal is based on mediated RSA that can be used to guarantee the integrity and authenticity of Web resources. The main idea behind this scheme (and any other IBC solution) is to generate and use public keys based on publicly available information that can be used to identify users or resources. On the other hand, private keys are generated by the Trusted Authority (TA) who possesses a master secret key (Section 3.1). Then, in Section 3.2, we explain what modifications (and simplifications) will be made in order to apply it for verifying Web resources.

3.1 Mediated RSA (mRSA)

One of the mRSA advantages is its *transparency*: in signature mode, mRSA yields standard RSA signatures which are much easier to incorporate with existing protocols. Mediated RSA involves a special entity called the SEM (SEcurity Mediator) which is a *partially* trusted server. To sign or decrypt a message, user Alice (one of the characters featuring in most cryptography scenarios) must obtain a message-specific token from the SEM.

The main idea behind mRSA is the splitting of an RSA private key into two parts using threshold cryptography. One part is given to the user while the other is given to the SEM. When the user and the SEM cooperate, the system is functionally equivalent to standard RSA. The fact that the private key is not held entirely by any one party is transparent to the outside, i.e. to those who use the corresponding public key to verify the signature (for more details see [3]).

3.2 Creating an Identity based Infrastructure for Resource Authenticity

Our proposal uses mRSA to address the problem of authenticity and integrity of Web resources. Consider a set of services offered by some organization and a set of recourses associated with each such service. Ideally, we would like any third party to be able to authenticate these resources without the use of public key infrastructures or complicated protocols.

The basic idea behind mRSA is the use of a single *common* RSA modulus N among all users of a system. In our case, however, the "users" are the services offered by the organization with resources tied to these services. These resources must be integrity protected and authenticated by anyone interacting with a particular service. Thus, these resources correspond to the "messages" that need to be signed and must bear the signature of the corresponding service.

Using the same modulus by multiple entities in a normal RSA setting is totally insecure since anyone, using its own knowledge of a single key-pair, can factor the modulus and compute the other entities' private keys. However, this does not apply in our setting since the private key is shared between the entity and the SEM. Thus an attacker must compromise *both* to undermine the security of the system.

In the following, we use the full name of a service as the unique identifier (public key) for that service. We use the notation $ID_{Service}$ to denote the identity that will be used to compute the public RSA exponent. During initialization, a trusted authority (TA) sets up the RSA modulus N for all the services of the organization. N is equal to the product of two large safe primes p and q. The public exponent $e_{Service}$ is the result of a hash function such as SHA1 on $ID_{Service}$, with the rightmost bit set to one so that with high probability $e_{Service}$ is relatively prime to $\phi(N)$. This process is shown below:

Generate Public Key for $ID_{Service}$

Let k be the security parameter (say $k = 2048$)

- Generate random $k/2$-bit primes r and s such that $p = 2r + 1$ and $q = 2s + 1$ are also primes.
- Set $N = pq$
- For a particular service identified by $ID_{Service}$
  1. Set $e_{Service} = \text{hash}(ID_{Service}) \parallel 1$
  2. Set $d = 1/e_{Service}\ mod\ \phi(N)$
  3. Set $d_{Service}$ equal to a random number in $Z_N - \{0\}$
  4. Set $d_{SEM} = d - d_{Service}\ mod\ \phi(N)$

Once the private key is generated for a particular service, it can be used to sign a resource R through collaboration with the SEM. In what follows, we assume the existence of an appropriate encoding scheme that can be used to break the multiplicative properties of RSA. Typically, one can use the Probabilistic Signature Scheme (PSS) for RSA that can be found in then Public-Key Cryptography family of Standards PKCS#1. RSA-PSS incorporates processing schemes designed to provide additional security for RSA signatures. This encoding scheme, although not shown in detail, should be used and is denoted by Hash-PSS in the following description:

Sign resource R

- Set h equal to Hash-PSS(R)
- Compute partial signatures PS_{SEM} and $PS_{Service}$ as follows:
  1. $PS_{SEM} = h^{d_{SEM}}\ mod\ N$
  2. $PS_{Service} = h^{d_{Service}}\ mod\ N$
- Set $S = PS_{SEM} \cdot PS_{Service}\ mod\ N$
- Return signature S

Finally, any interested party that wants to ensure the authenticity of the resource R, it can do so by first computing the public key $e_{Service}$ from available information and then verifying the signature S.

Verify Signature S on resource R

- Retrieve domain modulus N
- Set $e_{Service} = \text{hash}(ID_{Service}) \parallel 1$
- Compute $h = S^{e_{Service}}\ mod\ N$
- Verify whether h is equal to Hash-PSS(R)

Security issues

The security of this scheme depends on whether someone can break into the SEM and the server and retrieve the corresponding private keys. In general, this

is a safe assumption to make since the SEM can reside in hardened server that is more resistant to break-ins than usual machines.

This also solves the problem of the common modulus since even if a server is compromised, no attacker can use the key $d_{Service}$ to sign resources without the collaboration of the SEM. Additionally, knowing $d_{Service}$ for a particular service does not leak any information about either the primes that constitute the modulus or the private keys of other services. This is because both keys d_{SEM} and $d_{Service}$ are random quantities. Using a simulation argument, one can show that any attack that takes advantage of one of the two keys could be turned into an attack to standard RSA [3]. Thus knowledge of one of these keys does enable the attacker to sign fake resources (details omitted due to space restrictions).

4 IBC over the existing Web Infrastructure

4.1 IBC over the existing Web protocols

In our proposal, IBC is to be used to verify Web resources that can be identified by their original URL or URI. This is achieved by applying the IBC scheme of Section 3.2, using the URL/URI (or the DNS name portion of it) as part of the key and by using the XMLDsig standard format and processes. In this way, checking the authenticity and integrity of a resource can be more efficient in comparison to PKI-based schemes as the URL/URI of the obtained resource is well known and the *modulus* N for the domain of the resource can be known or promptly obtained from a secure server or DNSSEC [5].

Although the role of the URL is to provide the location of resources instead of identification, we assume that when a resource is not identified by a URI, its URL serves as its identifier. We make a distinction between an *administrative domain* and a *DNS name*. An administrative domain can manage one or more DNS names. The *DNS name portion of a URL/URI* is the *host* field of the *authority* component of a URL or URI [16].

The digital signature for a resource can indicate whether the signature was produced using the whole of the URL/URI as key or just the DNS name in it, depending on the policy of the domain from which the resource originates. Effectively, our proposal is for two different modes of IBC-based resource validation:

- **MODE 1:** The 'modulus N' of the domain is used in combination with the *DNS name* part of the resource URL/URI to sign it. This means that the same private key can be used to sign any document in a specific domain regardless of its URL/URI. This can be flexible in terms of private key and digital signature management. On the other hand, there is a higher risk in using a single private key to sign all domain resources.
- **MODE 2:** The 'modulus N' of the domain is used in combination with the whole URL/URI to sign it. This requires a different private key for each URL/URI in a domain. This makes the management of private keys and digital signatures in a domain more complex but is ideal in cases when URIs represent user identities, such as OpenIDs.

Our proposal requires no changes to existing protocols and infrastructure, only some extra functionality on the client (e.g. Web browser) side, which can be implemented as a client plug-in. The *KeyInfo* element of a XMLDsig signature can be used to indicate the mode of validation the client is expected to use (Mode 1 or Mode 2).

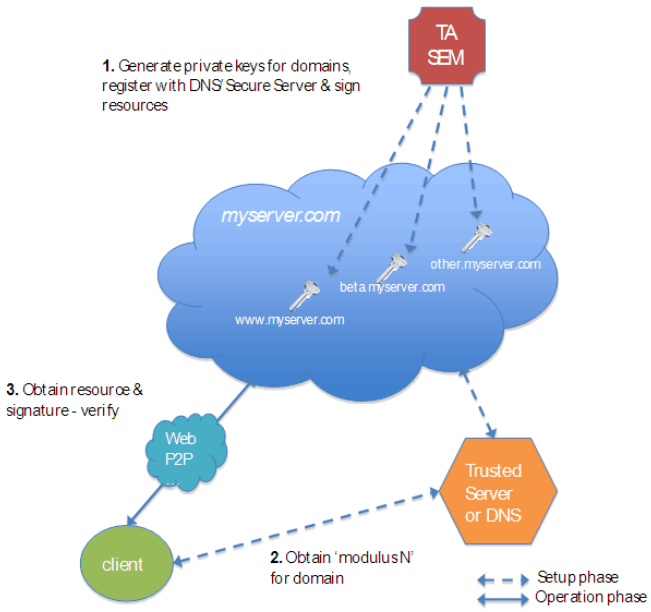

Fig. 1. Domain-wide IBC deployment for verifying Web resources.

Figure 1 shows an example of an administrative domain *myserver.com*, which runs three different servers with three different DNS names (*www.myserver.com*, *betamyserver.com* and *other.myserver.com*). The TA of the administrative domain *myserver.com* will provide the *modulus* N for all three DNS names and the SEM of *myserver.com* will sign each resource as detailed in Section 3.2. A client (Web browser, P2P client or other) can obtain and cache the 'modulus N' for any of the three DNS domains that can appear in URL/URIs, using a trusted server or secure DNS. When a client obtains a signed resource from these domains it will be able to verify its authenticity and integrity.

In order to support authenticity and integrity checks for non URL-based resources, the use of IBC for URI-identified resources can be implemented. Unlike a URL, a URI does not necessarily correspond to a communication end-point from which a resource can be obtained – it can be just an identifier. However, both URLs and URIs are expected to be maintained by the domain they belong to. For this reason, resources that claim to be identified by a URI can be checked for their authenticity and integrity by a browser, program or user by using the 'modulus N' of the domain name portion of the URI. In this way, it would

be possible to obtain any resource identifiable by a URI/URL via any type of content dissemination infrastructure and still be able to verify its authenticity and integrity. The 'modulus N' can be obtained in a number of ways as discussed next in Section 4.2.

The proposed two modes of resource validation can cater for different requirements of authenticity and integrity checks and support both domain-signed resources (Mode 1) or resources signed by individuals (Mode 2 for OpenID).

4.2 Scenarios for IBC deployment on the Web

The following scenarios are envisaged for the deployment of IBC based authentication and integrity of Web resources that can be identified by a URL or URI. In these scenarios, the content identified by a URL or URI can be obtained over a number of different content delivery channels, not necessarily the Web. However, the Web infrastructure is used to obtain the 'modulus N' for the domains of each URL/URI.

Scenario 1: IBC for URL-based or URI-identified resources ('modulus N' maintained per DNS domain)

In this scenario the 'modulus N' for the domain can be obtained by the client:

- From a Web server on the domain of the specific URL/URI. This requires the client to issue an HTTP GET request for a *standard relative URL* on the domain. For example, 'modulus N' for URL/URI of domain *www.myserver.com* could be obtained by issuing a GET request for the reserved relative URL '*modulusN.xml*' to the server of the domain, with absolute URL: *http://www.myserver.com/modulusN.xml*.
- If infrastructural changes for DNSSEC are adopted, 'modulus N' for a domain could be obtained by the DNSSEC [5] or by alternative directory services [8].

In all the above approaches, a client will obtain the 'modulus N' for the corresponding URL/URI domain and, when the URL-based resource is retrieved by the specific URL, will isolate the XMLDsig and proceed as detailed in Section 4.1. This approach is scalable and has the advantage that it can be easily deployed without necessarily making infrastructural changes. On the other hand, this scheme may have to rely on using a *reserved relative URL* on every domain (e.g. 'modulusN') and an agreed XML schema for 'modulus N' distribution (e.g. for file '*modulusN.xml*' in the example above).

Scenario 2: IBC for URL-based or URI identified resources ('modulus N' obtained from dedicated secure server)

This scenario applies when multiple DNS domains are shared within an organization or a virtual community. In this case, we assume that a dedicated secure server can be employed for the distribution of the 'modulus N' for URLs and

URIs available for the participating domains. The client software (or browser plug-ins) can be configured to contact the designated secure server to obtain 'modulus N' when necessary. This approach has the benefit that it does not require a separate 'modulus N' for each DNS domain but it is not scalable and may require manual configuration by the user.

5 Conclusions and Further Work

In this work we described a new approach for authenticity of web resources. Our proposal has a number of advantages ranging from transparency, ease of use, and implicit authentication of resources to seamless introduction and support of context-centric networking and collaboration in virtual communities. Rather than using certificate based Public Key Cryptography (PKI), our proposal is based on the use of the more intuitive Identity Based Cryptography. Once resources are signed using a version of RSA called mediated RSA, anybody can verify the authenticity of these resources simply by using the name (URL/URI) of the resource as the verification key. In our proposal, we still need to use a type of domain certificate that includes the common RSA modulus N, but we should stress that this "certificate" is not like a normal public key certificate but rather a *long lived attribute* certificate for the entire domain that can be retrieved either by a dedicated server or by using DNSSEC if infrastructural changes are adopted (Section 4.2).

One other difference with traditional PKI systems is that the private key is generated by the trusted authority (TA). This enforcement, in general, may raise concerns related to key escrow and privacy surrounding the management of private keys. The first concern is not really an issue in our case since we are only dealing with signature (correspondingly client verification) of resources, so no encryption takes place. For the second concern which may lead to compromise of private keys and signature non-repudiation once a server has been compromised, one could use multiple servers and threshold cryptography. Furthermore, the use of the SEM ensures that an attacker must compromise *both* to undermine the security of the system.

Our approach opens some interesting directions for research. We plan to investigate performance issues when signing content of different types and sizes. This is an issue that needs to be addressed since most of the web traffic increase over the last few years has been attributed to the exchange of large volume multimedia content. We also plan to identify further requirements to support collaborative authoring of resources in virtual communities and investigate the use of OpenID to let authors sign portions of collaboratively produced documents, which, in turn, could be double signed by the domain of the community. The management of resources, private keys and signatures to support both modes can be a challenging task that we aim to explore further.

References

1. Boneh, D.; Franklin, M. Identity-based encryption from the Weil pairing. In J. Kilian, editor, Advances in Cryptology - CRYPTO 2001, volume 2139 of LNCS, pages 213-229. Springer-Verlag, (2001)
2. Boneh, D.; Ding, X; Tsudik, G; M. Wong. Method for Fast Revocation of Public Key Certificates and Security Capabilities. In 10th USENIX Security Symposium, pp. 297-308 (2001)
3. Boneh, D.; Ding, X; Tsudik. Identity-Based Mediated RSA. In Proceedings of 3rd International Workshop on Information and Security Applications, WISA '02 (2002)
4. Crampton, J., Lim, H. W., and Paterson, K. G. What can identity-based cryptography offer to web services? In Proceedings of the 2007 ACM Workshop on Secure Web Services, ACM, New York, NY, 26-36 (2007)
5. R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. DNS Security Introduction and Requirements. IETF draft: draft-ietf-dnsext-dnssec-intro-13, October 10 (2004)
6. Gentry, C.; Silverberg, A. Hierarchical ID-Based Cryptography.In Proceedings of Advances in Cryptology, Asiacrypt '02, Lecture Notes in Computer Science, Springer-Verlag (2002)
7. J. Horwitz and B. Lynn. Toward hierarchical identity-based encryption.In Proc. EUROCRYPT 02, pages 466-481. Springer-Verlag, (2002)
8. Jones, J.P.; Berger, D.F.; Ravishankar, C.V. Layering public key distribution over secure DNS using authenticated delegation. 21st Annual Computer Security Applications Conference, (2005)
9. Metz C. and Bsales J. Five Ideas That Will Reinvent Modern Computing. PC Magazine, (2007)
10. O'Reilly, T. 2005. What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. O'Reilly (2005)
11. PARC (Palo Alto Research Center). Content-Centric Networking: PARC's Strategy for Pioneering a Self-Organizing Network That Meets Information Needs. Media Backgrounder. (2006) http://www.parc.com/content/newsroom/CCN_backgrounder.pdf
12. Paterson, K.G.; Price, G. A comparison between traditional public key infrastructures and identity-based cryptography. Information Security Technical Report, Volume 8, Issue 3, Pages 57-72, July (2003)
13. Shamir. Identity-based cryptosystems and signature schemes. In Advances in Cryptology - CRYPTO '84, volume 196 of LNCS, pages 47-53. Springer-Verlag (1984)
14. XMLDsig - XML Signature Syntax and Processing (2nd Edition). Eastlake, D., Reagle, J., Solo, D., Hirsch, F. and Roessler, T. W3C Recommendation, (2008) http://www.w3.org/TR/2008/REC-xmldsig-core-20080610/
15. RFC2807. XML Signature Requirements. IETF, July (2000) www.ietf.org/rfc/rfc2807.txt
16. RFC3986. Uniform Resource Identifier (URI): Generic Syntax. IETF, January (2005) www.ietf.org/rfc/rfc3986.txt
17. Smetters, D.K.; Durfee G. Domain-Based Administration of Identity-Based Cryptosystems for Secure Email and IPsec. In Proceedings of 12th USENIX Security Symposium, pp 215-229 (2003)