Design of a Security Mechanism for RESTful Web Service Communication through Mobile Clients

Femke De Backere, Brecht Hanssens, Ruben Heynssens, Rein Houthooft, Alexander Zuliani, Stijn Verstichel, Bart Dhoedt and Filip De Turck
Information Technology Department (INTEC), Ghent University - iMinds, Gaston Crommenlaan 8, bus 201, 9050 Ghent, Belgium
Email: Femke.DeBackere@intec.UGent.be

Abstract—Security is not taken into account by default in the Representational State Transfer (REST) architecture, but its layered architecture provides many opportunities for implementing it. In this paper, a security mechanism for Web Service communication through mobile clients devices is proposed, that conforms to the REST architecture as much as possible. This approach has been inspired by some known security mechanisms, but implemented in such a way that it focusses on statelessness and aims to be lightweight. Results indicate that the custom security mechanism outperforms the Transport Layered Security (TLS) based system. Because of the genericness of REST, the proposed security mechanism can be adopted by a wide variety of other RESTful Web Services.

I. INTRODUCTION

Representational State Transfer (REST) [1] is an architectural pattern, specifically tailored to building applications and Web Services that are distributed over the public Internet. Resources within the REST architecture can be manipulated through a set of unique Uniform Resource Identifiers (URIs). By simply sending universal Hypertext Transfer Protocol (HTTP) methods to these URIs, certain actions on the resources they represent can be performed. HTTP status codes are then used to send feedback to the client. These actions are already provided by HTTP, accelerating the adoption of REST [2]. The use of these URIs to transfer important data may result in privacy breaches as data is not anonymized.

As Web Services conforming to the REST constraints are based on HTTP, they suffer from the same inconveniences as many other standard web applications. The following malicious activities need to be taken into account when designing a good security mechanism:

- Unvalidated input: if no validation mechanism is required to check input data, adversaries can attack the backend server by making use of SQL injection. SQL commands are thus injected from e.g. a web form into the database of an application to access database content.
- Immature protocols or mechanisms: weaknesses in the used protocols or mechanisms can be exploited to attack both client and server. When making use of older protocols or outdated security mechanisms, attackers can relatively easily gain access to client or user content.
- Man-in-the-Middle (MITM) attack: an attacker makes independent connections with the victims and relays messages between them, making users believe that they are talking directly to each other, when in fact the entire conversation is controlled by the attacker.
- Replay attack: a valid data transmission is maliciously repeated. This attack is carried out either by the originator of the transmission or by an adversary who intercepts the data and retransmits it. When replaying the username and password, attackers can steal the identity of users.
- Spoofing: an attack, where a person or program successfully masquerades as another by falsifying data. Systems where the source of the data does not have to be authorized suffer from these attacks, making it possible to deliver falsified data to users.
- Message altering: an adversary changes the contents of a message that is being transmitted between client and server, without being noticed.
- Cross-Site Scripting (XSS) & Cross-Site Request Forgery (CSRF): both make use of a web application to transport an attack to the browser of the end user. This makes it possible to steal the authentication token or spoof content to the user.

RESTful Web Services are stateless. This means that every request is only dependent on itself, one simply has to examine the request to gather all the details concerning it. Stateless also means that the service is more reliable, because there are less steps where something can go wrong. In distributed services, the lack of state means that there is no overhead to keep the different servers consistent. A downside, however, is that data tends to be sent in a rather repetitive way, because no history is saved, possibly resulting in a larger overhead than typical stateful alternatives. The notion of state in larger applications is often very present, making communication in a stateless way a non-trivial task. Although there are some well-built RESTful Web Services, such as the Amazon Web Services (AWS) [3], still many REST Web Services fail to conform to the REST architecture guidelines. Many of these misuses have to do with the difficulty of implementing a good URI structure, or the indistinctness of mapping Web Service methods on the general HTTP methods.

REST is often treated as an alternative to Simple Object Access Protocol (SOAP), although both cannot be directly compared. REST is an architecture for building (web) applications, whereas SOAP is a protocol for exchanging structured
information between services and applications. SOAP is commonly used in today’s Web Services, and uses Web Services Security (WSS) [4] to cope with the aspect of security. SOAP allows the invocation of Remote Procedure Calls (RPCs) through HTTP ports, which prevents firewall configuration to moderate Web Services. Since it is no longer possible at the Transmission Control Protocol (TCP) level to distinguish the application, one has to analyze the packets at the HTTP level to comprehend their purpose. Many developers consider this a major flaw that compromises network security [5]. The Extensible Markup Language (XML) envelope within SOAP, consisting of a header and a body message, has to be inspected to understand the true intent of the message. In contrast, the intent of messages in REST can easily be analyzed by looking at the low weight HTTP methods used in the request.

This paper introduces a security mechanism, using only a bare minimum of non-RESTful elements, keeping the lightweight character of mobile clients in mind, which means preserving battery power, limiting data transfer and message overhead. The suggested implementation is then compared to a fully TLS-based solution.

The remainder of this paper is organized as follows. Related work of existing security mechanisms can be found in Section II. Section III will explain the custom security mechanism, followed by evaluation results in Section IV. Finally, conclusions can be found in Section V.

II. RELATED WORK

In Table I, a comparison is made between known security mechanisms, keeping in mind their suitability in a RESTful architecture. Nowadays, the capabilities and restrictions of smartphones need to be taken into account. Security mechanisms thus need to require low processing power to support the battery lifetime and the ability to function in varying network circumstances with respect to mobile interaction.

As RESTful Web Services are stateless, they do not usually have any kind of session, in which to perform a challenge-response mechanism. Popular known mechanisms such as Open Authorization (OAuth) [6] and OpenID [7] therefore violate the strict principles of a RESTful architecture, simply because they are not stateless. On the one hand, OAuth is an open standard for authorization that relies on sessions, in which client information must be stored at the server-side in order to authorize them. On the other hand, OpenID is an open standard for brokered authentication, which is able to authenticate a user by means of URI, but relies on the fact that this user needs to remember this URI. This is not stateless, and therefore not RESTful compatible. Storing these URIs might also lead to privacy breaches if logs are not protected and anonymized [11].

Transport Layer Security/Secure Sockets Layer (TLS/SSL) provides secure peer-to-peer authentication, but this mechanism is inadequate when requests for authentication are based on delegation, allowing sites to authenticate on behalf of their users [8]. HTTP Secure (HTTPS) is widely used for confidentiality, but it only provides hop-to-hop security [11]. An advantage is that clients can store the server certificate to avoid MITM attacks. It is also very easy to set up as it is a highly standardized mechanism. The downside of TLS/SSL is the high level of overhead used and the fact that it is very CPU intensive, because the connection link has to be created at every request, due to the stateless REST principle, which in turn heavily reduces scalability. However, most systems do make use of at least one connection with TLS, mostly during the login phase of the security mechanism. Datagram TLS (DTLS) [12] provides communication privacy for datagram protocols and can also be used to secure RESTful Web Services, if the overhead of the DTLS handshake is reduced [13]. However, by reducing the overhead, some RESTful principles are harmed, such as the usage of sessions.

A good solution, obeying the RESTful principles, is a token-based approach [8], [9], [3]. This approach lets the service generate a token on the first request, when users enter their username and password. The token is then sent to the client, who will add it to every further request to access a certain REST resource. This token should not be bound to any data, as it is merely a substitution for a username and password [9]. Once a token has been obtained the user can offer this token to the remote site, which guarantees them access to a specific REST resource for a certain amount of time. When the token is obtained by a malicious client, it can take over communication with the RESTful Web Service. Amazon Web Services (AWS) [3] also uses a token-based approach. AWS is seen by many as a well-built, well-secured RESTful application [11]. Their approach requires the client to add authentication data to each request. This data is based on a shared key, which is used to construct a hash-based message authentication code (HMAC). The data is sent to the server with a personal signature. This signature is then checked by AWS, who keeps a database mapping keys to IDs. The main advantage of this approach is that there are no user credentials sent by the client. It is also very lightweight since HMACs can be processed very fast. However, the disadvantage is that there is no proof that one possesses the secret key [5]. AWS avoids replay attacks by also adding timestamps to this mechanism [11]. Malisetti et al. proposed another custom token-based approach to cope with the security in REST [8].

The fundamental challenge with existing security models is that they offer IP-to-IP security solutions and not Application-to-Application ones. Therefore, there is a huge need to extend the existing HTTP security models to cope with RESTful Web Services [5].

Serme et al. presented a security model based on confidentiality and digital signatures to protect RESTful messages [2]. These messages carry tokens with them for non-repudiation and provide data confidentiality by encrypting its content. They propose a set of HTTP-headers to transmit meta-data, unlike WSS, which modifies messages to add its own container. They also make use of digital signatures to demonstrate the authenticity of messages, based on a ‘digest then encrypt’ process. Encryption is added to prevent attackers from being able to read the content of messages whilst staying unnoticed.
Peng et al. proposed an extended HTTP Digest Authentication Scheme (DAS) algorithm based on WSS [10]. This approach adds a secondary password to the original DAS algorithm. It can also resist to offline password guessing attacks as opposed to DAS, whilst having the same speed. Replay attacks can be avoided by adding a random nonce to every request.

The work by Story et al. is based on the use of Friend of a Friend (FOAF) and Secure Socket Layer (SSL) [14]. With this combination they make it possible for a server to authenticate a client given a simple URI. This URI can then be used directly for authorization or to explore more information in the web making use of linked data. FOAF is a Resource Description Framework (RDF) vocabulary, allowing each person to describe himself and his network of friends by linking oneself to acquaintances. Compared with Public Key Infrastructure (PKI), which also makes use of certificates, their approach removes the need for hierarchical authorities to assert identity. This makes it much more flexible and hence very suited in a distributed environment.

### III. Proposed Security Mechanism

In most existing mechanisms, a lot of the security elements used are stateful, and thus not conform to the RESTful principles. In the presented system, only the most essential non-RESTful elements are added. Whenever a non-RESTful element is added, a motivation is given concerning the reason why. The constructed system makes use of different cryptographic functions.

#### A. Security requirements

In many applications, four different types of client authentication for REST resource access can be distinguished: i) public, ii) proof of previous login is required, iii) direct login details are necessary and iv) an offline challenge-response system is recommended.

#### B. Login mechanism

The login mechanism used in the designed system is shown in Algorithm 1. Users authorize themselves by the combination of a username (un) and a password (pw). The back-end server will store this un and a hashed version of the pw combined with a salt: H(H(pw)+salt). By only sending and storing a hashed version of the pw, the server can never leak the passwords of users, even in case of a hack. Adding a salt prevents an adversary (in case of a cracked database) from performing an offline brute force or dictionary attack on different passwords simultaneously. This approach also greatly decreases the effectiveness of rainbow tables to reverse hashes.

The server should thus store at least the following data: un, H(H(pw)+salt), salt. Although storing user-bound data is not conform to the REST principles, it is essential for a strong security system.

The server authorizes the user by using a digital certificate. When downloading the application, the certificate is downloaded as well. This certificate is signed by a certificate authority (CA), ensuring its correctness. It can be checked by verifying the certificate chain up to the root CA. When
the client connects to the server, he/she will provide his/her un and pw. This combination cannot be sent in plain text for security reasons and the server has to be trusted first. That is why during the login phase, a TLS connection will be used. The client sends un/H(pw), this way his/her password never leaves the client system in plain text. By using a TLS connection, replay and offline guessing attacks can be avoided.

Upon receiving the combination of (un, H(pw)), the server will calculate H(H(pw)+salt) and compare it to the value stored in the database. After this, the server will calculate an authentication token AT and a symmetric key SK. It also binds an expiry date expDate to this token. The AT can be seen as RESTful, it is simply a surrogate for a un and pw. The SK and the expDate however are not. Either way, the SK is essential for providing data integrity and the expDate is needed to close the session when a user forgets to logout or his/her AT/ SK gets stolen. Now, the database holds the following data: un, H(H(pw)+salt), salt, SK, AT, expDate. When the expDate is reached, the AT/ SK is destroyed. This prevents other people to keep using this AT/ SK combination when using a public device where a previous user forgot to logout. After this the AT/ SK is sent to the client over the TLS connection, after which the connection is shut down.

C. REST resource access

The user-resource interaction mechanism used in the system is shown in Algorithm 2. When a client wishes to access a REST resource, he will add a timestamp TS and his AT to his request. The total message will be signed by a HMAC using the SK. A Hash-based Message Authentication Code (HMAC) provides message integrity and authentication. Because the user never has to sent his un again, this provides an extra layer of security. The AT will be thrown away after a time expDate, thus it is not as critical as a permanent username. By making use of an AT, the un and pw do not have to be stored on the (possibly public) client system.

On receiving this message, the server will look up the correct database entry by using the AT. Then, he will calculate the HMAC using the SK stored. He will also check if the expDate has not been reached and that the TS of the latest message is higher than the previous one. Because of the use of a TS, replay attacks are impossible. It thus ensures message freshness. However, using a user-bound timestamp is not RESTful. An adversary can never forge a request as he does not possess the SK to calculate the HMAC. This provides protection to offline guessing attacks. A request can also never be modified as a hash function is being used.

The server will add a TS to its own response to the client to ensure message freshness. The server can either encrypt the data by using the SK with symmetric encryption (or make use of TLS), or simply not encrypt the data at all. Either way, the server should authenticate its message by signing it with his own private key (the one used to sign the digital certificate). Alternatively, it can use the same HMAC as the client to authenticate itself. Upon acquisition of this

Algorithm 1: User login

<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>user fills in details and presses login button</td>
<td>user is logged in</td>
</tr>
</tbody>
</table>

1. $C \leftrightarrow S$: setupTLSConnection(serverCertificate)
2. $C$: TS$_C$ $\leftarrow$ getCurrentTime()
3. $C$: hash$_C$ $\leftarrow$ H(pw)
4. $C \rightarrow S$: sendLoginDetails(un, hash$_C$, TS$_C$)
5. $S$: hash$_S$, salt $\leftarrow$ lookupInDB(un)
6. $S$: verify(H(hash$_S_+$ salt), hash$_S$)
7. $S$: prevTS$_C$ $\leftarrow$ TS$_C$
8. $S$: AT $\leftarrow$ calcAuthToken()
9. $S$: SK $\leftarrow$ calcSecretKey()
10. $S$: TS$_S$ $\leftarrow$ getCurrentTime()
11. $S$: saveExpDateInDB(TS$_S$ + T)
12. $C \leftarrow S$: send(AT, SK, TS$_S$
13. $C$: prevTS$_S$ $\leftarrow$ TS$_S$

Algorithm 2: User-resource interaction

<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>user is logged in, possesses SK and AT</td>
<td>user receives resource representation</td>
</tr>
</tbody>
</table>

1. $C$: HTTPReq $\leftarrow$ createHTTPReq()
2. $C$: TS$_C$ $\leftarrow$ getCurrentTime()
3. $C$: HTTPReq.add(AT, TS$_C$
4. $C$: HMAC $\leftarrow$ HMAC(HTTPReq.body, SK)
5. $C$: HTTPReq.add(HMAC)
6. $C$: symEncrypt(HTTPReq.body, SK)
7. $C \rightarrow S$: send(HTTPReq)
8. $S$: AT $\leftarrow$ HTTPReq.get(AT)
9. $S$: expDate, SK, prevTS$_C$, $\leftarrow$ lookupInDB(AT)
10. $S$: symDecrypt(HTTPReq.body, SK)
11. $S$: verify(HTTPReq.TS$_C$ $>$ prevTS$_C$
12. $S$: verify(currentTime < expDate)
13. $S$: HMAC$_C$ $\leftarrow$ remove(HTTPReq.HMAC)
14. $S$: HMAC$_C$ $\leftarrow$ HMAC(HTTPReq.body, SK)
15. $S$: verify(HMAC$_C$ = HMAC$_C$
16. $S$: prevTS$_C$ $\leftarrow$ HTTPReq.TS$_C$
17. $S$: HTTPResp $\leftarrow$ createHTTPResp()
18. $S$: resource $\leftarrow$ retrieveResource()
19. $S$: TS$_S$ $\leftarrow$ getCurrentTime()
20. $S$: HTTPResp.add(resource, TS$_S$
21. $S$: sign $\leftarrow$ digSign(HTTPResp.body)
22. $S$: HTTPResp.add(sign)
23. $S$: symEncrypt(HTTPReq.body, SK)
24. $C \leftarrow S$: send(HTTPResp)
25. $C$: symDecrypt(HTTPReq.body, SK)
26. $C$: verify(HTTPResp.TS$_S$ $>$ prevTS$_S$
27. $C$: sign $\leftarrow$ remove(HTTPResp.sign)
28. $C$: verify(HTTPResp.body, sign, PK$_S$
29. $C$: prevTS$_S$ $\leftarrow$ HTTPResp.TS$_S$
response the client will check the TS to the last saved TS. If it is higher, the message passes the first check. Then the digital signature is checked using the certificate stored in the application. If this passes the test, the message is considered safe and thus accepted. Alternatively, instead of timestamps, nonces can be used. However, this requires extra message overhead as these nonces have to be sent back and forth at every request/response. Nevertheless, it does eliminate the requirement of clock synchronization between client and server, which is essential for loose coupling. When logging out, the user does not send a TS, only the SK is needed. When logged out the same thing happens as if the expDate was reached. The AT/SK are destroyed at server-side. If a user logs in on a different system, the old AT/SK combination is also destroyed. This message may be replayed as a logout or possible once. If a user forgets to log out, the next user of the (possibly public) device can only capture the AT/SK. The un/pw are not stored in the application. This AT/SK are only usable for a limited period of time. However, much damage can be done in this short period of time. This is why critical functions require the use of the un/H(pw) of the user.

If data confidentiality is required, it is possible to add symmetric encryption of the HTTPReq.body (see Algorithm 2). This can be done based on the SK between line 6 and 7 (client-side), then a decryption step needs to be added between lines 7 and 8 (server-side). The HTTPResp.body would be encrypted between lines 19 and 20 (server-side) and decrypted between lines 20 and 21 (client-side). However, encryption is not always needed for some applications as a lot of the data exchanged is not confidential. The symmetric encryption process on lines 6, 10, 23 and 25 in Algorithm 2 is optional and should only be used when data confidentiality is required. It can be switched on and off dynamically based on the communication context.

Denial-of-Service (DoS) attacks should be countered by making use of Google’s firewall filter. Spoofing is not possible since the adversary does not have the SK and does not know the username or the password. The weakest link in our system is the password of a certain user, it should therefore be long and hard to guess. Password checking mechanisms could be used to enforce the use of a strong password. It should be noted that a corrupted end-user system is unprotected. This cannot be fixed by security mechanisms on the system. For very critical actions, an offline challenge/response system should be used, much like the system used for bank transactions.

V. CONCLUSIONS

The aspect of security remains a big issue for RESTful Web Services. Many of the current security mechanisms violate the RESTful principles and are, because of their stateful nature, not able to cope with the scalability advantages that REST provides. Basic RESTful security standards, like HTTP authentication, are outdated and therefore omit vital security information from the application. This can be done in this short period of time. This is why critical functions require the use of the un/H(pw) of the user.

If data confidentiality is required, it is possible to add symmetric encryption of the HTTPReq.body (see Algorithm 2). This can be done based on the SK between line 6 and 7 (client-side), then a decryption step needs to be added between lines 7 and 8 (server-side). The HTTPResp.body would be encrypted between lines 19 and 20 (server-side) and decrypted between lines 20 and 21 (client-side). However, encryption is not always needed for some applications as a lot of the data exchanged is not confidential. The symmetric encryption process on lines 6, 10, 23 and 25 in Algorithm 2 is optional and should only be used when data confidentiality is required. It can be switched on and off dynamically based on the communication context.

Denial-of-Service (DoS) attacks should be countered by making use of Google’s firewall filter. Spoofing is not possible since the adversary does not have the SK and does not know the username or the password. The weakest link in our system is the password of a certain user, it should therefore be long and hard to guess. Password checking mechanisms could be used to enforce the use of a strong password. It should be noted that a corrupted end-user system is unprotected. This cannot be fixed by security mechanisms on the system. For very critical actions, an offline challenge/response system should be used, much like the system used for bank transactions.

IV. EXPERIMENTAL SETUP AND RESULTS

To evaluate the proposed security mechanism, it is applied to the login procedure of a prototype of a location-aware social network. Then, it is compared with a system that is fully TLS-based. This TLS-based system simply requires a user to login after which an authentication token is returned, which is added to every message. After this, every REST resource access is tunneled through a TLS link. According to Reese, this is a strong way to secure RESTful Web Services [15].

The two mechanisms are compared based on their message execution overhead in Figure 1. The client used for these measurements was a 2.66 GHz Intel Core 2 Duo with 4 GB RAM laptop. The test application was deployed on Google App Engine (GAE). For the messaging tests Wireshark has been used to compare the size of the IP-packets. The results indicate that the custom security system performs 36.4% better in terms of messaging overhead and 28.6% better in terms of processing overhead. Also the ratio of the standard deviation to the average value is lower for the custom built system (22.6%) compared to the TLS-based system (27.9%).

The scalability of the proposed mechanism was also investigated. The Virtual Wall at iMinds was used in combination with the Spirent Avalanche framework [16] to simulate a large number of users simultaneously. For the acquisition of this data, the request rate was steadily increased until the desired level and waited until the response time, averaged over a four second interval, was stabilized. The actual response time average and standard deviation were calculated over a 40-second interval. Figure 2 shows a large increase in deviation as the load increases, this is caused by the development server provided by Google, which is not meant for production use and as such it does not fairly schedule a large amount of simultaneous requests. The custom system started failing (dropping a high number of requests) around 100 requests per second. The TLS-based system lasted until around 140 requests per second. On GAE, this would not happen because of the automatic scaling provided for both the application and the database. These results remain relevant because it gives an idea of how both mechanisms compare under stress in terms of computational overhead. It is obvious the custom-built system contains less unnecessary overhead data. The system is also faster than the TLS system, because TLS has to perform a handshake and generate keys for every request.

solutions. TLS seems to be a usable standard, nonetheless, the overhead introduced is simply too large for non-continuous connections, as with mobile interaction. Therefore, a custom security mechanism is proposed, using only a bare minimum of non-RESTful elements. Comparing this implementation with a fully TLS-based solution shows that this method outperforms the latter, both on the aspect of messaging as processing overhead. Scalability benchmarks have revealed that a large number of simultaneous requests significantly influence the performance. Because of the genericness of REST, our proposed security mechanism can be adopted by a wide variety of other RESTful Web Services.

REFERENCES