



# Bulwark: Holistic and Verified Security Monitoring of Web Protocols

Lorenzo Veronese<sup>1,2(✉)</sup>, Stefano Calzavara<sup>1</sup>, and Luca Compagna<sup>2</sup>

<sup>1</sup> Università Ca' Foscari Venezia, Venezia, Italy  
lorenzo.veronese@tuwien.ac.at

<sup>2</sup> SAP Labs France, Mougins, France

**Abstract.** Modern web applications often rely on third-party services to provide their functionality to users. The secure integration of these services is a non-trivial task, as shown by the large number of attacks against Single Sign On and Cashier-as-a-Service protocols. In this paper we present Bulwark, a new automatic tool which generates formally verified security monitors from applied pi-calculus specifications of web protocols. The security monitors generated by Bulwark offer holistic protection, since they can be readily deployed both at the client side and at the server side, thus ensuring full visibility of the attack surface against web protocols. We evaluate the effectiveness of Bulwark by testing it against a pool of vulnerable web applications that use the OAuth 2.0 protocol or integrate the PayPal payment system.

**Keywords:** Formal methods · Web security · Web protocols

## 1 Introduction

Modern web applications often rely on third-party services to provide their functionality to users. The trend of integrating an increasing number of these services has turned traditional web applications into *multi-party* web apps (MPWAs, for short) with at least three communicating actors. In a typical MPWA, a *Relying Party* (RP) integrates services provided by a *Trusted Third Party* (TTP). Users interact with the RP and the TTP through a *User Agent* (UA), which is normally a standard web browser executing a *web protocol*. For example, many RPs authenticate users through the Single Sign On (SSO) protocols offered by TTPs like Facebook, Google or Twitter, and use Cashier-as-a-Service (CaaS) protocols provided by payment gateway services such as PayPal and Stripe.

Unfortunately, previous research showed that the secure integration of third-party services is a non-trivial task [4, 12, 18, 20–23]. Vulnerabilities might arise due to errors in the protocol specification [4, 12], incorrect implementation practices at the RP [18, 21, 22] and subtle bugs in the integration APIs provided by the TTP [23]. To secure MPWAs, researchers proposed different approaches,

---

L. Veronese—Now at TU Wien.

most notably based on *run time monitoring* [7, 10, 15, 16, 25]. The key idea of these proposals is to automatically generate security monitors allowing only the web protocol runs which comply with the expected, ideal run. Security monitors can block or try to automatically fix protocol runs which deviate from the expected outcome.

In this paper, we take a retrospective look at the design of previous proposals for the security monitoring of web protocols and identify important limitations in the current state of the art. In particular, we observe that:

1. existing proposals make strong assumptions about the placement of security monitors, by requiring them to be deployed either at the client [7, 15] or at the RP [10, 16, 25]. In our work we show that both choices are sub-optimal, as they cannot prevent all the vulnerabilities identified so far in popular web protocols (see Sect. 4);
2. most existing proposals are not designed with formal verification in mind. They can ensure that web protocol runs are compliant with the expected run, e.g., derived from the network traces collected in an unattacked setting, however they do not provide any guarantee about the actual security properties supported by the expected run itself [10, 16, 25].

Based on these observations, we claim that none of the existing solutions provides a reliable and comprehensive framework for the security monitoring of web protocols in MPWAs.

*Contributions.* In this paper, we contribute as follows:

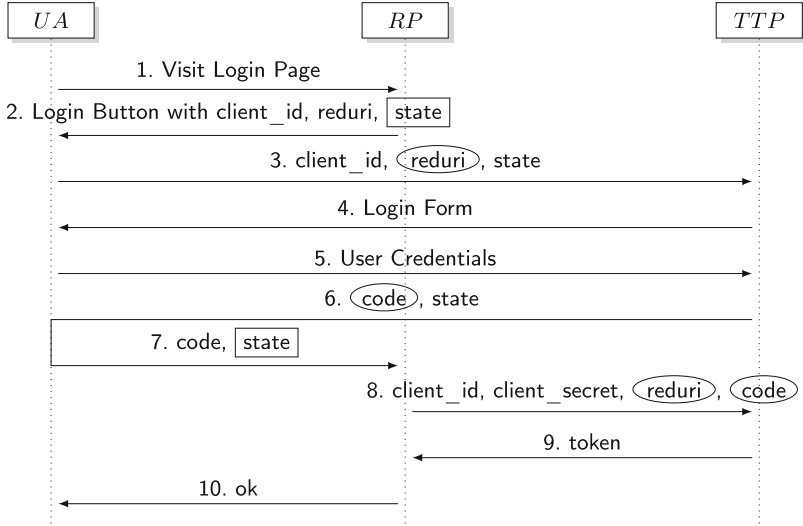
1. we perform a systematic, comprehensive design space analysis of previous work and we identify concrete shortcomings in all existing proposals by considering the popular OAuth 2.0 protocol and the PayPal payment system as running examples (Sect. 4);
2. we present Bulwark, a novel proposal exploring a different point of the design space. Bulwark generates formally verified security monitors from applied pi-calculus specifications of web protocols and lends itself to the appropriate placement of such monitors to have full visibility of the attack surface, while using modern web technologies to support an easy deployment. This way, Bulwark reconciles formal verification with practical security (Sect. 5);
3. we evaluate the effectiveness of Bulwark by testing it against a pool of vulnerable web applications that use the OAuth 2.0 protocol or integrate the PayPal payment system. Our analysis shows that Bulwark is able to successfully mitigate attacks on both the client and the server side (Sect. 6).

## 2 Motivating Example

As motivating example, shown in Fig. 1, we selected a widely used web protocol, namely OAuth 2.0 in explicit mode, which allows a RP to leverage a TTP for authenticating a user operating a UA.<sup>1</sup> The protocol starts (step 1) with the UA

---

<sup>1</sup> The protocol in the figure closely follows the Facebook implementation; details might slightly vary for different TTPs.



**Fig. 1.** Motivating example: Facebook OAuth 2.0 explicit mode

visiting the RP’s login page. A login button is provided back that, when clicked, triggers a request to the TTP (steps 2–3). Such a request comprises: `client_id`, the identifier registered for the RP at the TTP; `reduri`, the URI at RP to which the TTP will redirect the UA after access has been granted; and `state`, a freshly generated value used by the RP to maintain session binding with the UA. The UA authenticates with the TTP (steps 4–5), which in turn redirects the UA to the `reduri` at RP with a freshly generated value `code` and the `state` value (steps 6–7). The RP verifies the validity of `code` in a back channel exchange with the TTP (steps 8–9): the TTP acknowledges the validity of `code` by sending back a freshly generated `token` indicating the UA has been authenticated. Finally, the RP confirms the successful authentication to the UA (step 10).

Securely implementing such a protocol is far from easy and many vulnerabilities have been reported in the past. We discuss below two representative attacks with severe security implications.

**Session Swapping** [21]. Session swapping exploits the lack of contextual binding between the login endpoint (step 2) and the callback endpoint (step 7). This is often the case in RPs that do not provide a `state` parameter or do not strictly validate it. The attack starts with the attacker signing in to the TTP and obtaining a valid `code` (step 6). The attacker then tricks an honest user, through CSRF, to send the attacker’s `code` to the RP, which makes the victim’s UA authenticate at the RP with the attacker’s identity. From there on, the attacker can track the activity of the victim at the RP. The RP can prevent this attack by checking that the value of `state` at step 7 matches the one that was generated at step 2. The boxed shapes around `state` represent this invariant in Fig. 1.

**Unauthorized Login by Code Redirection** [4, 14]. Code (and token) redirection attacks exploit the lack of strict validation of the `reduri` parameter and involve its manipulation by the attacker. The attacker crafts a malicious page which fools the victim into starting the protocol flow at step 3 with valid `client_id` and `state` from an honest RP, but with a `reduri` that points to the attacker’s site. The victim then authenticates at the TTP and is redirected to the attacker’s site with the `code` value. The attacker can then craft the request at step 7 with the victim’s `code` to obtain the victim’s `token` (step 9) and authenticate as her at the honest RP. The TTP can prevent this attack by (i) binding the `code` generated at step 6 to the `reduri` received at step 3, and (ii) checking, at step 8, that the received `code` is correctly bound to the supplied `reduri`. The rounded shapes represent this invariant in Fig. 1.

### 3 Related Work

We review here existing approaches to the security monitoring of web protocols, focusing in particular on their adoption in modern MPWAs. Each approach can be classified based on the placement of the proposed defense, i.e., we discriminate between client-side and server-side approaches.

We highlight here that none of the proposed approaches is able to protect the entire attack surface of web protocols. Moreover, none of the proposed solutions, with the notable exception of WPSE [7], is designed with formal verification in mind and provides clear, precise guarantees about the actual security properties satisfied by the enforced policy.

#### 3.1 Client-Side Defenses

WPSE [7] extends the browser with a security monitor for web protocols that enforces the intended protocol flow, as well as the confidentiality and the integrity of messages. This monitor is able to mitigate many vulnerabilities found in the literature. The authors, however, acknowledge that some classes of attack cannot be prevented by WPSE. In particular, network attacks (like the HTTP variant of the IdP mix-up attack [12]), attacks that do not deviate from the intended protocol flow (like the automatic login CSRF from [4]) and purely server-side attacks are out of scope.

OAuthGuard [15] is a browser extension that aims to prevent five types of attacks on OAuth 2.0 and OpenID Connect, including CSRF and impersonation attacks. OAuthGuard essentially shares the same limitations of WPSE, due to the same partial visibility of the attack surface (the client side).

Recently Google has shown interest in extending its Chrome browser to monitor SSO protocols,<sup>2</sup> however their solution deals with a specific attack against their own implementation of SAML and is not a general approach designed to protect other protocols or TTPs.

<sup>2</sup> <https://gsuiteupdates.googleblog.com/2018/04/more-secure-sign-in-chrome.html>.

### 3.2 Server-Side Defenses

InteGuard [25] focuses on the server side of the RP, as its code appears to be more error-prone than that of the TTP. InteGuard is deployed as a proxy in front of the RP that checks invariants within the HTTP messages before they reach the web server. Different types of invariants are automatically generated from the network traces of SSO and CaaS protocols and enable the monitor to link together multiple HTTP sessions in transactions. Thanks to its placement, InteGuard can also monitor back channels (cf. steps 8–9 of Fig. 1). The authors explicitly mention that InteGuard does not offer protection on the TTP, expecting further efforts to be made in that direction. Unfortunately, several attacks can only be detected at the TTP, e.g., some variants of the unauthorized login by auth. code (or token) redirection attack from [4].

AEGIS [10] synthesizes run time monitors to enforce control-flow and data-flow integrity, authorization policies and constraints in web applications. The monitors are server-side proxies generated by extracting invariants from a set of input traces. AEGIS was designed for traditional two-party web applications, hence it does not offer comprehensive protection in the multi-party setting, e.g., due to its inability to monitor messages exchanged on the back channels. However, as mentioned by the authors, it can still mitigate those vulnerabilities which can be detected on the front channel of the RP (e.g., the shop-for-free TomatoCart example [10]). Similar considerations apply to BLOCK [16], a black-box approach for detecting *state violation* attacks, i.e., attacks which exploit logic flaws in the application to allow some functionality to be accessed at inappropriate states.

Guha et al. [13] apply a static control-flow analysis for JavaScript code to construct a *request-graph*, a model of a well-behaved client as seen by the server application. They then use this model to build a reverse proxy that blocks the requests that violate the expected control flow of the application, and are thus marked as potential attacks. Also this approach was designed for two-party web applications, hence does not offer holistic protection in the multi-party setting. Moreover, protection can only be enforced on web applications which are entirely developed in JavaScript.

## 4 Design Space Analysis

Starting from our analysis of related work, we analyze the design space of security monitors for web protocols, discussing pros and cons along different axes. Our take-away message is that solutions which assume a fixed placement of a single security monitor, which is the path taken by previous work, are inherently limited in their design for several reasons.

### 4.1 Methodology

We consider three possible deployment options for security monitors: the first two are taken from the literature, while the last one is a novel proposal we make. In particular, we focus on:

1. *browser extensions* [7,15]: a browser extension is a plugin module, typically written in JavaScript, that extends the web browser with custom code with powerful capabilities on the browser internals, e.g., arbitrarily accessing the cookie jar and monitoring network traffic;
2. *server-side proxies* [10,16,25]: a proxy server acts as an intermediary sitting between the web server hosting (part of) the web application and the clients that want to access it;
3. *service workers*: the Service Worker API<sup>3</sup> is a new browser functionality that enables websites to define JavaScript workers running in the background, decoupled from the web application logic. Service workers provide useful features normally not available to JavaScript, e.g., inspecting HTTP requests before they leave the browser, hence are an intriguing deployment choice for client-side security monitors.

We evaluate these options with respect to four axes, originally proposed as effective criteria for the analysis of web security solutions [8]:

1. *ease of deployment*: the practicality of a large-scale deployment of the defense mechanism, i.e., the price to pay for site operators to grant security benefits;
2. *usability*: the impact on the end-user experience, e.g., whether the user is forced to take actions to activate the protection mechanism;
3. *protection*: the effectiveness of the defense mechanism, i.e., the supported and unsupported security policies;
4. *compatibility*: the precision of the defense mechanism, i.e., potential false positives and breakages coming from security enforcement.

## 4.2 Ease of Deployment and Usability

Service workers are appealing, since they score best with respect to ease of deployment and usability. Specifically, the deployment of a service worker requires site operators to just add a JavaScript file to the web application: when a user visits the web application, the installation of the service worker is transparently performed with no user interaction.

Server-side proxies similarly have the advantage of ensuring transparent protection to end users. However, they are harder to deploy than service workers, as they require site operators to have control over the server networking. Even if site operators just needed to apply small modifications to the monitored application, they would have to reroute the inbound/outbound traffic to the proxy. This is typically easy for the TTP, which is usually a major company with full control over its deployment, but it can be impossible for some RPs. RPs are sometimes deployed on managed hosting platforms that may not allow any modification on the server itself, except for the application code. Note that site operators could implement the logic of the proxy directly in the application code, but this solution is impractical, since it would require a significant rewriting of the web

<sup>3</sup> [https://developer.mozilla.org/en-US/docs/Web/API/Service\\_Worker\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Service_Worker_API).

application logic. This is also particularly complicated when the web application is built on top of multiple programming languages and frameworks.

Finally, browser extensions are certainly the worst solution with respect to usability and ease of deployment. Though installing a browser extension is straightforward, site operators cannot assume that every user will perform this manual installation step. In principle, site operators could require the installation of the browser extension to access their web application, but this would have a major impact on usability and could drive users away. Users’ trust in browser extensions is another problem on its own: extensions, once installed and granted permissions, have very powerful capabilities on the browser internals. Moreover, the extension should be developed for the plethora of popular browsers which are used nowadays. Though major browsers now share the same extension architecture, many implementation details are different and would force developers to release multiple versions of their extensions, which complicates deployment. In the end, installing an extension is feasible for a single user, but relying on browser extensions for a large-scale security enforcement is unrealistic.

### 4.3 Protection and Compatibility

We study protection and compatibility together, since the *visibility* of the attack surface is the key enabler of both protection and compatibility. Indeed, the more the monitor has visibility of the protocol messages, the more it becomes able to avoid both false positives and false negatives in detecting potential attacks.

We use the notation  $P_1 \leftrightarrow P_2$  to indicate the channel between two parties  $P_1$  and  $P_2$ . If a monitor has visibility over a channel, then the monitor has visibility over all the messages exchanged on that channel.

**Visibility.** Browser extensions run on the UA and can have visibility of all the messages channeled through it. In particular, browser extensions can request *host permissions* upon installation to get access to the traffic exchanged with arbitrary hosts, which potentially enables them to inspect and edit any HTTP request and response relayed through the UA. In MPWAs, the UA can thus have visibility over both the channels  $UA \leftrightarrow RP$  and  $UA \leftrightarrow TTP$  (shortly indicated as  $UA \leftrightarrow \{RP, TTP\}$ ). However, the UA itself is not in the position to observe the entire attack surface against web protocols: for example, when messages are sent on the back channel between the RP and the TTP ( $RP \leftrightarrow TTP$ ) like in our motivating example (steps 8–9 in Fig. 1), an extension is unable to provide any protection, as the UA is not involved in the communication at all.

Server-side proxies can be categorized into *reverse proxies* and *forward proxies*, depending on whether they monitor incoming or outgoing HTTP requests respectively (plus their corresponding HTTP responses). Both approaches are useful and have been proposed in the literature, e.g., InteGuard [25] uses both a reverse proxy and a forward proxy at the RP to capture messages from the UA and to the TTP respectively. This way, InteGuard has full visibility of all the messages flowing through the RP, i.e.,  $RP \leftrightarrow \{UA, TTP\}$ . However, this is still

not sufficient to fully monitor the attack surface. In particular, server-side proxies cannot inspect values that never leave the UA, like the *fragment identifier*, which instead plays an important role in the implicit flow of OAuth 2.0.

Finally, web applications can register service workers at the UA by means of JavaScript. Service workers can act as network proxies with access to all the traffic exchanged between the UA and the origin<sup>4</sup> which registered them. This way, service workers have the same visibility of a reverse-proxy sitting at the server; however, since they run on the UA, they also have access to values which never leave the client, like the fragment identifier. Despite this distinctive advantage, service workers are severely limited by the Same Origin Policy (SOP). In particular, they cannot monitor traffic exchanged between the UA and other origins, which makes them less powerful than browser extensions. For example, contrary to browser extensions like WPSE [7], a single service worker cannot monitor and defend both the RP and the TTP. This limitation can be mitigated by using multiple service workers and/or selectively relaxing SOP using CORS, which however requires collaboration between the RP and the TTP. Since service workers are more limited than browser extensions, they also share their inability to monitor back channels, hence they cannot be a substitute for forward proxies.

In the end, we conclude that browser extensions and server-side proxies are *complementary* in their ability to observe security-relevant protocol components, given their respective positioning, while service workers are strictly less powerful than their alternatives.

**Cross-Site Scripting (XSS).** XSS is a dangerous vulnerability which allows an attacker to inject malicious JavaScript code in a benign web application. Once a web application suffers from XSS, most confidentiality and integrity guarantees are lost, hence claiming security despite XSS is wishful thinking. Nevertheless, we discuss here how browser extensions can offer better mitigation than service workers in presence of XSS vulnerabilities. Specifically, since service workers can be removed by JavaScript, an attacker who was able to exploit an XSS vulnerability would also be able to void the protection offered by service workers. This can be mitigated by defensive programming practices, e.g., overriding the functions required for removing service workers, but it is difficult to assess both the correctness and the compatibility impact of this approach. For example, the deactivation of service workers might be part of the legitimate functionality of the web application or the XSS could be exploited before security-sensitive functions are overridden. Browser extensions, instead, cannot be removed by JavaScript and are potentially more robust against XSS. For example, WPSE [7] replaces secret values with random placeholders before they actually enter the DOM, so that secrets exchanged in the monitored protocol cannot be exfiltrated via XSS; the placeholders are then replaced with the intended values only when they leave the browser towards authorized parties.

---

<sup>4</sup> An origin is a triple including a scheme (HTTP, HTTPS, ...), a host ([www.foo.com](http://www.foo.com)) and a port (80, 443, ...). Origins represent the standard web security boundary.



**Table 1.** Attacks on OAuth 2.0 and PayPal

Attack		Channels to observe	UA		RP		TTP	
			<i>ext</i>	<i>sw</i>	<i>proxy</i>	<i>sw</i>	<i>proxy</i>	
<i>OAuth 2.0</i>								
1	307 Redirect attack [12]	UA ↔ TTP	✓	×	×	✓	✓	
2	Access token eavesdropping [21]	UA ↔ RP	✓	✓	✓	×	×	
3	Code/State Leakage via referer header [11, 12]	UA ↔ RP	✓	✓	✓	×	×	
4	Code/Token theft via XSS [21]	UA ↔ RP	✓	×	×	×	×	
5	Cross Social-Network Request Forgery [4]	UA ↔ RP	✓	✓	✓	×	×	
6	Facebook implicit AppId Spoofing [20, 23]	UA ↔ TTP	×	×	×	✓	✓	
7	Force/Automatic login CSRF [4, 21]	UA ↔ RP	✓	✓	✓	×	×	
8	IdP Mix-Up attack [12] (HTTP variant)	UA ↔ RP	×	×	✓	×	×	
9	IdP Mix-Up attack [12] (HTTPS variant)	UA ↔ RP	✓	✓	✓	×	×	
10	Naive session integrity attack [12]	UA ↔ RP	✓	✓	✓	×	×	
11	Open Redirector in OAuth 2.0 [14, 17]	UA ↔ {RP, TTP}	✓	✓	✓	✓	✓	
12	Resource Theft by Code/Token Redirection [4, 7]	UA ↔ TTP	✓	×	×	×	✓	
13	Session swapping [14, 21]	UA ↔ RP	✓	✓	✓	×	×	
14	Social login CSRF on stateless clients [4, 14]	UA ↔ RP	✓	✓	✓	×	×	
15	Social login CSRF through TTP Login CSRF [4]	UA ↔ TTP	✓	×	×	✓	✓	
16	Token replay implicit mode [14, 20, 24]	UA ↔ RP	✓	×	×	×	×	
17	Unauth. Login by Code Redirection [4, 14]	UA ↔ TTP	✓	×	×	×	✓	
<i>PayPal</i>								
18	<i>NopCommerce</i> gross change in IPN callback [22]	RP ↔ {UA, TTP}	×	×	✓	×	×	
19	<i>NopCommerce</i> gross change in PDT flow [22]	RP ↔ {UA, TTP}	×	×	✓	×	×	
20	Shop for free by malicious <i>PayeeId</i> replay [18, 20]	RP ↔ {UA, TTP}	×	×	✓	×	×	
21	Shop for less by <i>Token</i> replay [18, 20]	UA ↔ RP	×	×	✓	×	×	

**Tamper Resistance.** Since both browser extensions and service workers are installed on the client, they can be tampered with or uninstalled by malicious users or software. This means that the defensive practices put in place by browser extensions and service workers are voided when the client cannot be trusted. This is particularly important for applications like CaaS, where malicious users might be willing to abuse the payment system, e.g., to shop for free. Conversely, server-side proxies are resilient by design to this kind of attacks, since they cannot be accessed at the client side, hence they are more appropriate for web applications where the client cannot be trusted to any extent.

**Assessment on MPWAs.** We now substantiate our general claims by means of a list of known attacks against the OAuth 2.0 protocol and the PayPal payment system, two popular protocols in MPWAs. Table 1 shows this list of attacks. For each attack, we show which channels need to be visible to detect the attack and we conclude whether the attack can be prevented by a browser extension (*ext*), a service worker (*sw*) or a server-side proxy (*proxy*) deployed on either the RP or the TTP.

In general we can see that, in the OAuth 2.0 setting, a browser extension is the most powerful tool, as it can already detect and block by itself most of the attacks (15 out of 17). The exceptions are the Facebook implicit AppId Spoofing

attack [20, 23], which can only be detected at the TTP, and the HTTP variant of the IdP Mix-Up attack [12], which is a network attack not observable at the client. Yet, remarkably, a comparable amount of protection can be achieved by using just service workers alone: in particular, the use of service workers at both the RP and the TTP can stop 13 out of 17 attacks. The only notable differences over the browser extension approach are that: (i) the code/token theft via XSS cannot be prevented, though we already discussed that even browser extensions can only partially mitigate the dangers of XSS, and (ii) the resource theft by code/token redirection and the unauthorized login by auth. code redirection cannot be stopped, because they involve a cross-origin redirect that service workers cannot observe by SOP. Remarkably, the combination of service workers and server-side proxies offers transparent protection that goes beyond browser extensions alone: 16 out of 17 attacks are blocked, with the only exception of code/token theft via XSS as explained.

The PayPal setting shows a very different trend with respect to OAuth 2.0. Although it is possible to detect the attacks on the client side, it is not safe to do so because both browser extensions and service workers can be uninstalled by malicious customers. For example, such client-side approaches cannot prevent the shop for free attack of [18], where a malicious user replaces the merchant id with her own account id. Moreover, it is worth noticing that PayPal deliberately makes heavy use of back channels (RP  $\leftrightarrow$  TTP), since messages which are not relayed by the browser cannot be tampered with by malicious customers. This means that server-side proxies are the way to go to protect PayPal-like payment systems, as confirmed by the table.

#### 4.4 Take-Away Messages

Here we highlight the main take-away messages of our design space analysis. In general, we claim that different web protocols require different protection mechanisms, hence every defensive solution which is bound to a specific placement of monitors does not generalize. More specifically:

- A clear total order emerges on the ease of deployment and usability axes. Service workers score best there, closely followed by server-side proxies, whose deployment is still feasible and transparent to end-users. Browser extensions are much more problematic, especially for large-scale security enforcement.
- With respect to the protection and compatibility axes, browser extensions are indeed a powerful tool, yet they can be replaced by a combination of service workers and server-side proxies to enforce transparent protection, extended to attacks which are not visible at the client alone.

In the end, we argue that a combination of service workers and server-side proxies has the potential to reconcile security, compatibility, ease of deployment and usability. In our approach, described in the next section, we thus pursue this research direction.

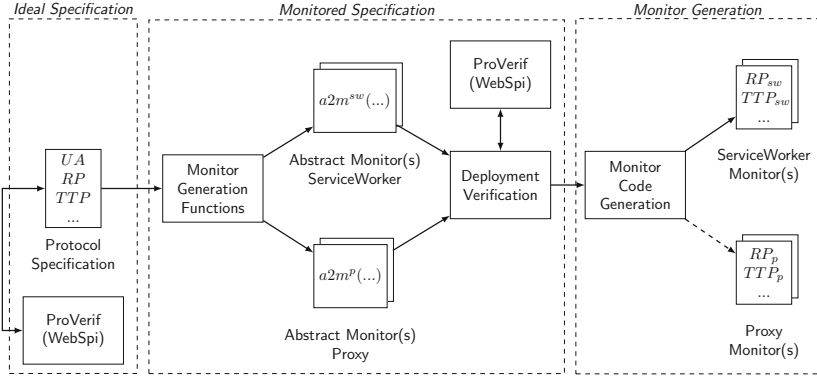


Fig. 2. Monitor generation pipeline

## 5 Proposed Approach: Bulwark

In this section we present Bulwark,<sup>5</sup> our formally verified approach to the holistic security monitoring of web protocols. For space reasons, we present an informal overview and we refer to the online technical report for additional details [2].

### 5.1 Overview

Bulwark builds on top of ProVerif, a state-of-the-art protocol verification tool [5]. ProVerif was originally designed for traditional cryptographic protocols, not for web protocols, but previous work showed how it can be extended to the web setting by using the WebSpi library [4]. In particular, WebSpi provides a ProVerif model of a standard web browser and includes support for important threats from the web security literature, e.g., *web attackers*, who lack traditional Dolev-Yao network capabilities and attack the protocol through a malicious website.

Bulwark starts from a ProVerif model of the web protocol to protect, called *ideal specification*, and generates formally verified security monitors deployed as service workers or server-side proxies. This process builds on an intermediate step called *monitored specification*. The workflow is summarized in Fig. 2.

To explain the intended use case of Bulwark, we focus on the typical setting of a multi-party web application including a TTP which offers integration to a set of RPs, yet the approach is general. The TTP starts by writing down its protocol in ProVerif, expressing the intended security properties by means of standard *correspondence assertions* (authentication) and (*syntactic*) *secrecy queries* supported by ProVerif. For example, the code/token redirection attack against OAuth 2.0 (cf. Sect. 2) can be discovered through the violation of a correspondence assertion [4]. The protocol can then be automatically verified for

<sup>5</sup> Bulwark is currently proprietary software at SAP: the tool could be made available upon request and an open-source license is under consideration.

security violations and the TTP can apply protocol fixes until ProVerif does not report any flaw. Since ProVerif is a sound verification tool [6], this process eventually leads to a security proof for an unbounded number of protocol sessions, up to the web model of WebSpi. The WebSpi model, although expressive, is not a complete model of the Web [4]. For example, it does not model advanced security headers such as `Content-Security-Policy`, frames and frame communication (`postMessage`). However, the library models enough components of the modern Web to be able to capture all the attacks of Table 1.

Once verification is done, the TTP can use Bulwark to automatically generate security monitors for its RPs from the ideal specification, e.g., to complement the traditional protocol SDK that the TTP normally offers anyway with protection for RPs, which are widely known to be the buggiest participants [25]. The TTP could also decide to use Bulwark to generate its own security monitors, so as to be protected even in the case of bugs in its own implementation.

## 5.2 Monitored Specification

In the *monitored specification* phase, Bulwark relaxes the ideal assumption that all protocol participants are implemented correctly. In particular, user-selected protocol participants are replaced by *inattentive* variants which comply with the protocol flow, but forget relevant security checks. Technically, this is done by replacing the ProVerif processes of the participants with new processes generated by automatically removing from the honest participants all the security checks (pattern matches, get/insert and conditionals) on the received messages, which include the invariants represented by the boxed checks in Fig. 1. This approximates the possible mistakes made by *honest-but-buggy* participants, obtaining processes that are interoperable with the other participants, but whose possible requests and responses are a super set of the original ones. Intuitively, an inattentive participant may be willing to install a monitor to prevent attackers from exploiting the lack of forgotten security checks. On the other hand, a deliberately malicious participant has no interest in doing so.

Then, Bulwark extracts from the ideal specification all the security invariant checks forgotten by the inattentive variants of the protocol participants and centralizes them within security monitors. This is done by applying two functions  $a2m^p$  and  $a2m^{sw}$ , which derive from the participant specifications new ProVerif processes encoding security monitors deployed as a server-side proxy or a service worker respectively. The  $a2m^p$  function is a modified version of the  $a2m$  function of [19], which generates security monitors for cryptographic protocols. The proxy interposes and relays messages from/to the monitored inattentive participant, after performing the intended security checks. A subtle point here is that the monitor needs to keep track of the values that are already in its knowledge and those which are generated by the monitored participant and become known only after receiving them. A security check can only be executed when all the required values are part of the monitor knowledge.

The  $a2m^{sw}$  function, instead, is defined on top of  $a2m^p$  and the ideal  $UA$  process. This recalls that a service worker is a client-side defense that acts as a

reverse proxy: a subset of the checks of both the server and the client side can be encoded into the final process running on the client. The function has three main responsibilities: (i) rewriting the proxy to be compatible with the service worker API; (ii) removing the channels and values that a service worker is not able to observe; (iii) plugging the security checks made by the ideal  $UA$  process into the service worker.

*Example 1.* Figure 3 illustrates how the  $RP$  of our OAuth 2.0 example from Sect. 2 is replaced by  $I(RP)$ , an inattentive variant in which the `state` parameter invariant is not checked. The right-hand side of the figure presents the  $RP$  as in Fig. 1 with a few more details on its internals, according to the ideal specification: (i) upon reception of message 1, the  $RP$  issues a new value for `state` and saves it together with the  $UA$  session cookie identifier (i.e., `state` is bound to the client session `sid(UA)`); and (ii) upon reception of message 7, the  $RP$  checks the `state` parameter and its binding to the client session. The inattentive version  $I(RP)$ , as generated by Bulwark, is shown on the left-hand side of Fig. 3: the `state` is neither saved by  $I(RP)$  nor checked afterward. The left-hand side of the figure also shows the proxy  $M(RP)$  generated by Bulwark as  $a2m^P(RP)$  to enforce the `state` parameter invariant at  $RP$ . We can see that the saving and the checking of `state` are performed by  $M(RP)$ . It is worth noticing that the  $M(RP)$  can only save `state` upon reception of message 2 from  $I(RP)$ . The service worker monitor  $a2m^{sw}(RP)$  would look very similar to  $M(RP)$ .

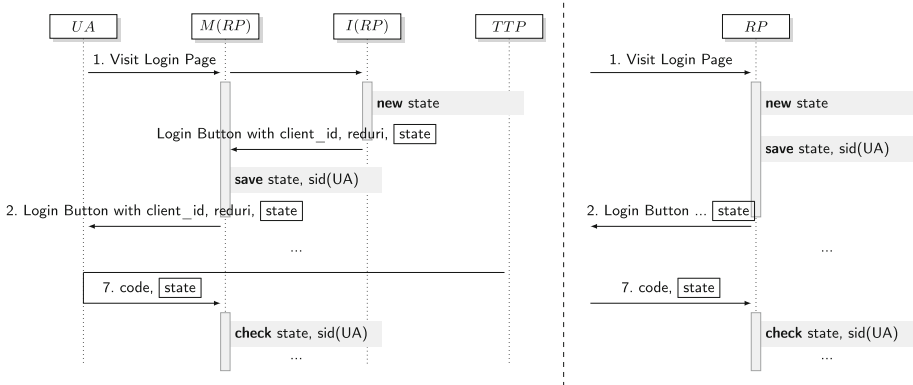


Fig. 3. Monitor invariant example

Finally, Bulwark produces a *monitored specification* where each inattentive protocol participant deploys a security monitor both at the client side (service worker) and at the server side (proxy). However, this might be overly conservative, e.g., a single service worker might already suffice for security. To optimize ease of deployment, Bulwark runs again ProVerif on the possible monitor deployment options, starting from the most convenient one, until it finds a setting which

satisfies the security properties of the ideal specification. As an example, consider the system in which the only inattentive participant is the  $RP$ . There are three possible options, in decreasing order of ease of deployment:

1.  $TTP \parallel I(RP) \parallel (a2m^{sw}(RP, UA) \parallel UA)$ , where the monitor is deployed as a service worker registered by the  $RP$  at the  $UA$ ;
2.  $TTP \parallel (I(RP) \parallel a2m^p(RP)) \parallel UA$ , where the monitor is a proxy at  $RP$ ;
3.  $TTP \parallel (I(RP) \parallel a2m^p(RP)) \parallel (a2m^{sw}(RP, UA) \parallel UA)$ , with both.

The first option which passes the ProVerif verification is chosen by Bulwark.

### 5.3 Monitor Generation

Finally, Bulwark translates the ProVerif monitor processes into real service workers (written in JavaScript) or proxies (written in Python), depending on their placement in the monitored specification. This is a relatively direct one-to-one translation, whose key challenge is mapping the ProVerif messages to the real HTTP messages exchanged in the web protocol. Specifically, different RPs integrating the same TTP will host the protocol at different URLs and each TTP might use different names for the same HTTP parameters or rely on different message encodings (JSON, XML, etc.).

Bulwark deals with this problem by means of a configuration file, which drives the monitor generation process by defining the concrete values of the symbols and data constructors that are used by the ProVerif model. When the generated monitor needs to apply e.g., a data destructor on a name, it searches the configuration file for its definition and calls the corresponding function that deconstructs the object into its components. Since data constructors/destructors are directly written in the target language as part of the monitor configuration, different implementations can be generated for the same monitor, so that a single monitor specification created for a real-world participant e.g., the Google TTP, can be easily ported to others, e.g., the Facebook TTP, just by tuning their configuration files.

## 6 Experimental Evaluation

### 6.1 Methodology

To show Bulwark at work, we focus on the core MPWA scenarios discussed in Sect. 4.3. We first write ideal specifications of the OAuth 2.0 explicit protocol and the PayPal payment system in ProVerif + WebSpi. We also define appropriate correspondence assertions and secrecy queries which rule out all the attacks in Table 1 and we apply known fixes until ProVerif is able to prove security for the ideal specifications. Then, we setup a set of case studies representative of the key vulnerabilities plaguing these scenarios (see Table 2). In particular, we selected vulnerabilities from Table 1 so as to evaluate Bulwark on both the RP and TTP via a combination of proxy and service worker monitors. For each case study, we

choose a set of inattentive participants and we collect network traces to define the Bulwark configuration files mapping ProVerif messages to actual protocol messages. Finally, we use Bulwark to generate appropriate security monitors and deploy them in our case studies. All our vulnerable case studies, their ideal specifications, and the executable monitors generated by Bulwark are provided as an open-source package to the community [1].

**Case Studies.** We consider a range of possibilities for OAuth 2.0. We start from an entirely artificial case study, where we develop both the RP and the TTP, introducing known vulnerabilities in both parties (CS1). We then consider integration scenarios with three major TTPs, i.e., Facebook, VK and Google, where we develop our own vulnerable RPs on top of public SDKs (CS2-CS4). Finally, we consider a case study where we have no control of any party, i.e., the integration between Overleaf and Google (CS5). We specifically choose this scenario, since the lack of the state parameter in the Overleaf implementation of OAuth 2.0 introduces known vulnerabilities.<sup>6</sup> To evaluate the CaaS scenario, we select legacy versions of three popular e-commerce platforms, suffering from known vulnerabilities in their integration with PayPal, in particular osCommerce 2.3.1 (CS6), NopCommerce 1.6 (CS7) and TomatoCart 1.1.15 (CS8).

**Evaluation Criteria.** We evaluate each case study in terms of four key aspects: (i) *security*: we experimentally confirm that the monitors stop the exploitation of the vulnerabilities; (ii) *compatibility*: we experimentally verify that the monitors do not break legitimate protocol runs; (iii) *portability*: we assess whether our ideal specifications can be used without significant changes across different case studies; and (iv) *performance*: we show that the time spent to verify the protocol and generate the monitors is acceptable for practical use.

**Table 2.** Test set of vulnerable applications

CS	RP	TTP	Protocol	Vuln. (Table 1)
1	<i>artificial RP 1</i>	<i>artificial IdP</i>	OAuth 2.0 explicit	#13 #17
2	<i>artificial RP 2</i>	facebook.com	OAuth 2.0 exp. (graph-sdk 5.7)	#13
3	<i>artificial RP 3</i>	vk.com	OAuth 2.0 exp. (vk-php-sdk 5.100)	#13
4	<i>artificial RP 4</i>	google.com	OAuth 2.0 exp. (google/apiclient 2.4)	#13
5	overleaf.com	google.com	OAuth 2.0 explicit	#13 #14
6	osCommerce 2.3.1	paypal.com	PayPal Standard	#18 #20
7	NopCommerce 1.6	paypal.com	PayPal Standard	#18
8	TomatoCart 1.1.15	paypal.com	PayPal Standard	#21

<sup>6</sup> We responsibly disclosed the issue to Overleaf and they fixed it before publication.

## 6.2 Experimental Results

The evaluation results are summarized in Table 3 and discussed below. In our case studies, we considered as inattentive participants all the possible sets of known-to-be vulnerable parties, leading to 10 experiments; when multiple experiments can be handled by a single run of Bulwark, their results are grouped together in the table, e.g., the experiments for CS2-CS4. Notice that for CS1 we considered three sets of inattentive participants: only TTP (vulnerability #17); only RP (vulnerability #13); and both RP and TTP (both vulnerabilities). Hence, we have 3 experiments for CS1, 3 experiments for CS2-CS4, 1 experiment for CS5 and 3 experiments for CS6-CS8.

**Table 3.** Generated monitors and run-time

CS	Ideal Spec.	Verification Time	Inattentive Parties	Monitor		Gen. Monitors				Prevented Vuln.
				Verification time	#verif.	RP	TTP	<i>sw proxy</i>	<i>sw proxy</i>	
1	IS1	29m	TTP	41m	2	×	×	×	✓	#17
			RP	15m	1	✓	×	×	×	#13
			RP,TTP	54m	3	✓	×	×	✓	#13 #17
2 3 4	IS1	27m	RP	18m	1	✓	×	×	×	#13
5	IS1*	19m	RP	17m	1	✓	×	×	×	#13 #14
6 7 8	IS2	3m	RP	8m	1	×	✓	×	×	#18 #20 #21

**Security and Compatibility.** To assess security and compatibility, we created manual tests to exploit each vulnerability of our case studies and we ran them with and without the Bulwark generated monitors. In all the experiments, we confirmed that the known vulnerabilities were prevented only when the monitors were deployed (security) and that we were able to complete legitimate protocol runs successfully both with and without the monitors (compatibility). Based on Table 3, we observe that 5 experiments can be secured by a service worker alone, 4 experiments can be protected by a server-side proxy and only one experiment needed the deployment of two monitors. This heterogeneity confirms the need of holistic security solutions for web protocols.

**Portability.** We can see that the ideal specification IS1 created for our first case study CS1 is portable to CS2-CS4 without any change. This means that different TTPs supporting the OAuth 2.0 explicit protocol like Facebook, VK and Google can use Bulwark straightaway, by just tuning the configuration file to their settings. This would allow them to protect their integration scenarios with RPs that (like ours) make use of the state parameter. This is interesting, since different TTPs typically vary on a range of subtle details, which are all accounted for correctly by the Bulwark configuration files. However, the state



parameter is not mandatory in the OAuth2 standard and thus TTPs tend to allow integration also with RPs that do not issue it. Case study CS5 captures this variant of OAuth 2.0: removing the state parameter from IS1 is sufficient to create a new ideal specification IS1\*, which enables Bulwark towards these scenarios as well. As to PayPal, the ideal specification IS2 is portable to all the case studies CS6-CS8. Overall, our experience indicates that once an ideal specification is created for a protocol, then it is straightforward to reuse it on other integration scenarios based on the same protocol.

**Performance.** We report both the time spent to verify the ideal specification (Verification Time) as well as the time needed to verify the monitors (Monitor Verification). Both steps are performed offline and just once, hence the times in the table are perfectly fine for practical adoption. Verifying the ideal specification never takes more than 30 min, while verifying the monitors might take longer, but never more than one hour in our experiments. The time spent in the latter step depends on how many runs of ProVerif are required to reach a secure monitored specification (see the very end of Sect. 5.2). For example, the first experiment runs ProVerif twice (cf. #verif.) and requires 41 min, while the second experiment runs ProVerif just once and thus takes only 15 min.

## 7 Conclusion

In this paper we identified shortcomings in previous work on the security monitoring of web protocols and proposed Bulwark, the first holistic and formally verified defensive solution in this research area. Bulwark combines state-of-the-art protocol verification tools (ProVerif) with modern web technologies (service workers) to reconcile formal verification with practical security. We showed that Bulwark can generate effective security monitors on different case studies based on the OAuth 2.0 protocol and the PayPal payment system.

As future work, we plan to extend Bulwark to add an additional protection layer, i.e., on client-side communication based on JavaScript and the postMessage API. This is important to support modern SDKs making heavy use of these technologies, like the latest versions of the PayPal SDKs, yet challenging given the complexity of sandboxing JavaScript code [3]. On the formal side, we would like to strengthen our definition of “inattentive” participant to cover additional vulnerabilities besides missing invariant checks. For example, we plan to cover participants who forget to include relevant security headers and are supported by appropriately configured monitors in this delicate task. Finally, we would like to further engineer Bulwark to make it easier to use for people who have no experience with ProVerif, e.g., by including support for a graphical notation which is compiled into ProVerif processes, similarly to the approach in [9].

**Acknowledgments.** Lorenzo Veronese was partially supported by the European Research Council (ERC) under the European Unions Horizon 2020 research (grant agreement No. 771527-BROWSEC).

## References

1. Bulwark case studies. <https://github.com/secgroup/bulwark-experiments>
2. Bulwark: holistic and verified security monitoring of web protocols (Technical report). <https://secgroup.github.io/bulwark-experiments/report.pdf>
3. Van Acker, S., Sabelfeld, A.: JavaScript sandboxing: isolating and restricting client-side JavaScript. In: Aldini, A., Lopez, J., Martinelli, F. (eds.) FOSAD 2015-2016. LNCS, vol. 9808, pp. 32–86. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-43005-8\\_2](https://doi.org/10.1007/978-3-319-43005-8_2)
4. Bansal, C., Bhargavan, K., Maffei, S.: Discovering concrete attacks on website authorization by formal analysis. In: CSF 2012. IEEE (2012)
5. Blanchet, B.: An efficient cryptographic protocol verifier based on prolog rules. In: CSFW 2001. IEEE (2001)
6. Blanchet, B.: Automatic verification of correspondences for security protocols. *J. Comput. Secur.* **17**(4), 363–434 (2009)
7. Calzavara, S., Focardi, R., Maffei, M., Schneidewind, C., Squarcina, M., Tempesta, M.: WPSE: fortifying web protocols via browser-side security monitoring. In: USENIX Security 18. USENIX Association (2018)
8. Calzavara, S., Focardi, R., Squarcina, M., Tempesta, M.: Surviving the web: a journey into web session security. *ACM Comput. Surv.* **50**(1), 1–34 (2017)
9. Carbone, R., Compagna, L., Panichella, A., Ponta, S.E.: Security threat identification and testing. In: ICST 2015. IEEE Computer Society (2015)
10. Compagna, L., dos Santos, D., Ponta, S., Ranise, S.: Aegis: automatic enforcement of security policies in workflow-driven web applications. In: CODASPY 2017. ACM (2017)
11. Fett, D., Küsters, R., Schmitz, G.: The web SSO standard OpenID connect: in-depth formal security analysis and security guidelines. In: CSF 2017. IEEE (2017)
12. Fett, D., Küsters, R., Schmitz, G.: A comprehensive formal security analysis of OAuth 2.0. CCS 2016. ACM (2016)
13. Guha, A., Krishnamurthi, S., Jim, T.: Using static analysis for ajax intrusion detection. In: WWW 2009. ACM (2009)
14. Hardt, D.: The OAuth 2.0 authorization framework. RFC 6749, October 2012
15. Li, W., Mitchell, C.J., Chen, T.: OAuthGuard: protecting user security and privacy with OAuth 2.0 and OpenID connect. In: SSR (2019)
16. Li, X., Xue, Y.: BLOCK: a black-bOx approach for detection of state violation attacks towards web applications. In: ACSAC 2011 (2011)
17. Lodderstedt, T., McGloin, M., Hunt, P.: OAuth 2.0 threat model and security considerations. RFC 6819, January 2013
18. Pellegrino, G., Balzarotti, D.: Toward black-box detection of logic flaws in web applications. In: NDSS (2014)
19. Pironi, A., Jürjens, J.: Formally-based black-box monitoring of security protocols. In: Massacci, F., Wallach, D., Zannone, N. (eds.) ESSoS 2010. LNCS, vol. 5965, pp. 79–95. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-11747-3\\_7](https://doi.org/10.1007/978-3-642-11747-3_7)
20. Sudhodanan, A., Armando, A., Carbone, R., Compagna, L.: Attack patterns for black-box security testing of multi-party web applications. In: NDSS (2016)
21. Sun, S.T., Beznosov, K.: The devil is in the (implementation) details: an empirical analysis of OAuth SSO systems. In: CCS 2012. ACM (2012)
22. Wang, R., Chen, S., Wang, X., Qadeer, S.: How to shop for free online - security analysis of cashier-as-a-service based web stores. In: S&P. IEEE (2011)

23. Wang, R., Chen, S., Wang, X.: Signing me onto your accounts through Facebook and Google: a traffic-guided security study of commercially deployed single-sign-on web services. In: S&P 2012. IEEE (2012)
24. Wang, R., Zhou, Y., Chen, S., Qadeer, S., Evans, D., Gurevich, Y.: Explicating SDKs: uncovering assumptions underlying secure authentication and authorization. In: USENIX (2013)
25. Xing, L., Chen, Y., Wang, X., Chen, S.: InteGuard: toward automatic protection of third-party web service integrations. In: NDSS 2013 (2013)