The Effect of HTML5 on Information Security

Jing AN*, Lei ZHANG, Xue-chao ZHANG and Chun-ian YOU

No 23, Taiping Road, Beijing, China
*Corresponding author

Keywords: HTML, Web application, Information security.

Abstract. As the new generation of Web language, HTML 5 has a lot of new attractive features, it can be said to be the huge progress of Web standards in nearly a decade. This paper focuses on its impact on Web applications and security problems on the basis of the development course of HTML5, and countermeasures are put forward briefly from two aspects of management and technology.

Introduction

Hypertext markup language (HTML) is a standard programming language for describing web content and appearance. HTML 5 is the new revision of the hypertext markup language, the goal is to replace 1999 HTML4.01 and XHTML 1.0 standards, in order to make the web standards to adapt to the rapid development of Internet applications, and meet the demand of actual network. HTML 5 is the greatest progress in the field of Web development standards in the last decade. It has many improvements on the basis of HTML 4, including: new enhanced flexibility of parsing rules, a new attribute, instant pattern painting, on-time media playing, etc., these improvements not only bring more interactivity to Internet browsers, also create new vulnerabilities for web security.

The Progress of Development

The first official version of HTML was HTML 2.0, launched by the Internet engineering task force (IETF). Subsequently, the world wide web consortium (W3C) replaced the Internet engineering task group (IETF), which took over the HTML standard. After 1990, the HTML version was frequently revised until 1999, when the last version of HTML, HTML 4.01, was created. By now, HTML has reached its first inflection point, and the Web world is beginning to change.

In 2000, the world wide web consortium (W3C) introduced the first revision of HTML 4.01 -- XHTML 1.0. X means "eXtensible"; Others interpret it as "eXtreme". The only difference between XHTML 1.0 and HTML 4.01 is that the syntax requires strict specification. If XHTML 1.0 is an xml-style HTML, XHTML 1.1 is a real XML. XHTML 1.1 was unable to use the text/HTML mime-type direct output, which only uses XML mime-type, while the prevailing browsers simply couldn't support the XML mime-type. The world wide Web consortium (W3C) urgently hope to bring Web to XML, and release the XHTML 2, XHTML 2 and XHTML 1 sound the same, but there are many different, XHTML 2 is not compatible with HTML, it completely become a whole new language. The creation of the world wide Web consortium (W3C) has caused a backlash among Web developers, with representatives from Opera, Apple and Mozilla beginning to voice their objections.

In 2004, Opera's Ian Hickson began to expand on HTML to adapt to new Web applications, which the world wide Web consortium (W3C) rejected. So they created the WHATWG, a research group for hypertext applications. The WHATWG research includes two parts: Web Forms 2.0 and Web Apps 1.0, they are HTML extensions, and eventually merge together to become today's HTML 5 working draft. While WHATWG is working on HTML 5, the W3C continues to make XHTML 2, but it is in a quandary.

In October 2006, Tim berners-lee, the father of the Web, published a blog that points out that the path from HTML to XML doesn't work. A few months later, the world wide web consortium
(W3C) formed a new HTML team, acknowledged the WHATWG research, as the foundation, on January 22, 2008, published its first official draft HTML 5. Nonetheless, the world wide web consortium (W3C) didn't end XHTML 2 until 2009, ending the chaotic scenes of XHTML 2 and HTML 5.

In May 2011, the world wide web consortium (W3C) announced that the development of the HTML 5 is nearing completion, work has started on the last hearing, if still no objections before May 22, the opinions of the existing HTML 5 specification, then the standard will become the international standard in 2012. Now it is already a standard, it is supported by the major browser vendors, such as Firefox, Chrome, Opera, Safari, Internet Explorer 9, some game developers and the mobile platform are in HTML 5 as a major development platform and the development direction of the future.

The Impact on Web Applications

In a broad sense, HTML 5 actually refers to a set of techniques, including HTML, CSS, and JavaScript. It can not only provide rich Internet application service and powerful information processing service, but also reduce the needing for browser plug-in (such as Adobe Flash, Microsoft Silverlight, etc.), and to provide more effective enhanced standard set of network application, all relevant Web application (including video, audio, images, animation, and interactions with the computer, etc.) based on HTML 5 can be standardized.

Specifically, HTML 5 adds a number of new syntactic features, including elements `<video>`, `<audio>`, and `<canvas>`, while integrating SVG. These elements are added in order to make it easier to add and process multimedia and image content in a web page. Other new elements `<section>`, `<article>`, `<header>`, and `<nav>` are enrich the data content of the document. New attributes are added for the same purpose. There are also some attributes and elements removed. Some elements like `<a>`, `<cite>` and `<menu>` are modified, redefined or normalized. At the same time, APIs and DOM have become the basics of HTML 5. HTML 5 also defines the details of processing illegal documents, allowing all browsers and client programs to handle syntax errors consistently. Unlike previous versions, HTML 5 is not just for Web content, its new mission is to bring the Web to a mature application platform.

The Web Services Mode

The application run through a Web browser instead of a specific operating system in HTML 5. Users will be able to access the same programs and cloud-based content without limitation of location and device, and implement "Web-oriented". No matter when and where, it can be smoothly to complete all the tasks in any device ,so it changes the habit of users, the network service mode, and the balance of power in the mobile telecommunications, media and other technology industry.

The Network Interactivity

The last version of HTML is relatively troublesome to process the mixed content of text, audio and video multimedia, it is needed to split between words, pictures, audio, video, parsing the corresponding URL and handled in a different way. HTML 5 is totally unrestricted in this aspect, it uses the Canvas object to load images, that can calculate and draw according to the data, supports regular editor such as picture movement, rotation, scaling and so on, and add the function of video tags, help Web developers to add video easily, flexible display all kinds of content, enhance network interactive.

The Mobile Application Development

Web Storage API of HTML 5 can be seen as an enhanced cookie, without being limited by the data size. It has better flexibility and architecture, so data can be written to the ROM of the machine, it can also open the restored data again after close the browser. The offline caching not only reduce the network traffic, also do not take up any background resources, and reduce the pressure of
equipment hardware. At the same time, a form element specifically tailored for mobile platforms is added to the HTML5, which lays the foundation for developing mobile applications.

**The Information Security**

Currently, Web applications are subject to both malicious advertising and unauthorized changes. HTML 5 improves on this security risk and plans to add sandbox properties to the iframe element. This improvement will help to protect against malicious third party ads or prevent the replaying of untrusted content, and improve network security. In addition, the advent of HTML 5 will reduce the need for plugins for Web applications, thereby eliminating the information security problems caused by plug-ins.

**The Network Development**

HTML 5 is a unified development environment, provides a unified development languages (JavaScript), uniform data model (XML and DOM) and a set of rules binding text, audio frequency, video and graphics (CSS). Under the HTML 5, it need not to switch between HTML 5 and other development environments. Greatly simplify network development, reduce r&d and maintenance cost.

**The Defense of Information Security**

The new functionality of HTML 5, which drives the progress of Web applications, also increases the risk that attackers will launch some form of attack, posing a threat to their own security.

**The Security of Client Stores**

In previous HTML, it was only allowed to store Cookies as local information and allocate relatively little space. Clients tend to store a small amount of information such as simple session ID, when the user needs to access the same data for many times, it needed to send the request to the server side for many times, as a result, the Web access performance is greatly reduced.

With the increasing complexity and volume of Web applications, access performance becomes an important key for development. For this reason, HTML 5 has been tweaked, and its LocalStorage allows the browser to store large amounts of data on the client side and allows for using new data types. While this adjustment greatly improves access performance, it is at the expense of security, and the security implications are enormous. Under such a mechanism, for instance, sensitive data will be stored on the client side, the attacker will be able to easily access to sensitive data by simple methods, such as physical access or damage the client. If the client is not safe, the attacker can login page, steal the user certificate by URL storage, SQL injection attacks, or producing fake e-commerce website.

**The Cross-domain Communication Security**

In previous HTML, there was a significant security restriction called the Same-Origin Policy. This policy makes a significant limitation on the page content that JavaScript code can access, which is that JavaScript can only access content that contains in the same document domain. For example, the JavaScript code contained in the page under baidu.com cannot access the content of the page under google.com. Even pages between different subdomains can't be accessed through JavaScript code. The effect of Ajax is that it can't submit requests to different domains through an Ajax request implemented by XMLHttpRequest. HTML 5 provides a cross-domain communication mechanism bypassing the cagnote policy to enable the sharing of cross-source resources. This mechanism will allow servers in different domains to communicate between the Web browser's iframe. So the attacker can abuse the feature to get sensitive data.

**The WebSocket Security**

The WebSocket protocol is one of the new HTML 5 protocols that implement a two-way communication between the browser and the server, which can save server resources and bandwidth
and achieve real-time communication. It is only allowed to interact with the site over HTTP in traditional browsers, that consumed a lot of server bandwidth and resources, and apparently unable to meet the Web applications such as chat and gaming. Therefore, HTML 5 provides the socket API, the browser and server only need one handshake to create a fast channel for data transfer. This is highly vulnerable to abuse by attackers, making secret backdoor communications.

The New Elements Security

The most obvious change in HTML 5 is the addition of new elements related to Web animation, video, and the properties of existing elements. The security of these elements is not verified, and there may be known or unknown security vulnerabilities. For example, image processing has always been a dangerous field of overflow vulnerability, and there is no guarantee that the `<canvas>` and other elements will not exist vulnerability at the implementation time. At the same time, the unknown security problem made the protective devices that using blacklist strategy do not know how to start, the attacker can directly attack by using the new tags and attributes bypass the blacklist strategy.

Summary

Faced with the security challenges posed by HTML 5, developers can address both management and technology:

First, it is needed to start from a management perspective to consider whether the HTML 5 standards are appropriate and whether there is a serious security threat. If it do needed to use the new HTML 5 standards after consideration, larger adjustments to existing network security must be made. For example, increasing firewall, upgrading network environment, and using a VPN or private connection between your branches to improve the server's trustworthiness.

Second, it is needed to take appropriate safety measures from a technical perspective. For example, adding the necessary data validation and strictly controlling the actions performed by the plug-in.

References
