


Article

Design of Enhanced Document HTML and the Reliable Electronic Document Distribution Service

Hyun-Cheon Hwang¹ and Woo-Je Kim^{2,*} ¹ Graduate School of Public Policy and Information Technology, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea; a.hwang@seoultech.ac.kr² Department of Industrial and Information System Engineering, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea

* Correspondence: wjkim@seoultech.ac.kr

Abstract: Electronic documents are becoming increasingly popular in various industries and sectors as they provide greater convenience and cost-efficiency than physical documents. PDF is a widely used format for creating and sharing electronic documents, while HTML is commonly used in mobile environments as the foundation for creating web pages displayed on mobile devices, such as smartphones and tablets. HTML is becoming a more critical document format as mobile environments have been raised as the primary communication channel nowadays. However, HTML does not have the standard content integrity feature, and an electronic document based on HTML consists of a set of related files. Therefore, it has a vulnerability in terms of reliable electronic documents. We have proposed Document HTML, a single independent file with extended meta tags, to be a reliable electronic document and Chained Document, a single independent file with a blockchain network to secure content integrity and delivery assurance. In this paper, we improved the definition of Document HTML and researched certified electronic document intermediaries. Additionally, we designed and validated the electronic document distribution service using Enhanced Document HTML for real usability. Moreover, we conducted experimental verification using a tax notification electronic document, which has one of the top distribution volumes in Korea, to confirm how Document HTML provides a content integrity verification feature. Document HTML can be used in an enterprise that must send a reliable electronic document to a customer with an electronic document delivery service provider.

Keywords: HTML; Document HTML; reliable electronic document

Citation: Hwang, H.-C.; Kim, W.-J. Design of Enhanced Document HTML and the Reliable Electronic Document Distribution Service. *Electronics* **2023**, *12*, 2176. <https://doi.org/10.3390/electronics12102176>

Academic Editors: Juan M. Corchado, Carlos A. Iglesias, Byung-Gyu Kim, Rashid Mehmood, Fuji Ren and In Lee

Received: 2 April 2023
Revised: 2 May 2023
Accepted: 5 May 2023
Published: 10 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A document is essential to communicate between an enterprise and a customer. It contains sensitive, personalized information that is exchanged in various formats and channels. In the digital era, electronic documents have replaced traditional physical documents. Electronic documents mimic the layout and formatting of traditional physical documents and are easily readable by humans. PDF (Portable Document Format) is the de-facto standard format for electronic documents as it can be viewed or printed like a physical document and has the ability to maintain content integrity through the use of digital signatures. However, the user experience on mobile devices can be uncomfortable due to their smaller screen size compared to paper. HTML-based electronic documents are more suitable for mobile environments, but they do not ensure content integrity like PDF-based electronic documents do. Despite HTML being the main format for mobile environments, content integrity is still at risk. HTML does not have resources embedded and relies on external resources to display content. Additionally, a standard specification is necessary to ensure content integrity. The proposed Document HTML includes digital signatures to create a reliable electronic document [1]. Chained Document is an extension of Document HTML using blockchain technology to ensure content integrity and delivery assurance [2].

However, these approaches may have weaknesses in using meta tag declarations in terms of usability and compatibility because the meta tag is not readable since it is based on the comment tag. In addition, there is a vulnerability when loading external resources through embedded CSS files. Additionally, these did not consider how reliable content is distributed in real-world scenarios by third-party electronic document delivery service providers. Therefore, in this paper, we needed to improve the design of Document HTML by introducing a new style of meta tag and conformance. Moreover, we investigated how Extended Document HTML works with certified document delivery services in the real world. We selected an actual electronic document and performed experimental verification of the Extended Document HTML. As a result, this paper proposes Extended Document HTML and reliable delivery services that can be used in the real world. We conducted related research for the electronic document area and the digital signature technology in Section 2 and improved the Document HTML and investigated third-party electronic document delivery service providers in Section 3. We designed an electronic document distribution system with Document HTML and certified document delivery services in Section 4 and experimental verification in Section 5. We discussed the limitations of this research in Section 6 and finally concluded the value of this research in Section 7.

2. Related Research

2.1. PDF

PostScript by Adobe was developed to generate a digital document for digital printing publishing, and PDF was developed based on PostScript to extend digital documents for both digital and printing channels [3]. It became the ISO standard in 2008. PDF is a cross-platform and device-independent format that can display documents as intended. The latest version is PDF 2.0, and there are sub-specifications for each purpose [4], as shown in Table 1.

Table 1. Comparison of characteristics by PDF sub-specification.

Type	Purpose	Key Characteristic
PDF/A	Long term archive	<ul style="list-style-type: none"> • Prohibit multi-media object • Prohibit dynamic script • Prohibit password encryption • Prohibit external resources linkage
PDF/X	Digital print	<ul style="list-style-type: none"> • Require output intent • Embed fonts • Embed image objects
PDF/UA	Accessibility for visually impaired	<ul style="list-style-type: none"> • All objects must be tagged • Proper order of tags • Alternative text as a tag for non-text object

PDF is the de-facto standard for an electronic document as it can present content with a layout similar to physical paper and is an independent format. However, electronic documents based on PDF have poor readability in mobile environments as mobile device screens are generally smaller than physical paper in general. Electronic documents based on PDF give an uncomfortable user experience in the mobile environment because of poor readability on mobile devices. In addition, an uncomfortable user experience is getting worse as it does not have interactive features. Because of these reasons, electronic documents based on HTML are a more suitable format for mobile environments as it provides responsive content presentation and interactive features to help navigate content.

2.2. HTML

HTML was invented to share scientific documents on the web and create a structured document that a computer system can read. The first version of HTML on W3C (World Wide Web Consortium) is HTML 3.2, and the latest version is HTML 5.3. HTML uses a tag

to describe the structure of the HTML document, and there is content in the tag [5]. It was a typical way to create an HTML document without CSS (Cascading Style Sheet) under the HTML 3 specification, and the <table> tag is used widely to represent the document layout. HTML 4 uses the <div> tag and CSS to split layout and content, and CSS has been enhanced until now. However, the content layout using the <div> tag gives an unclear document structure as the <div> tag does not have a semantic meaning. Therefore, semantic tags, such as <section>, <header>, and <footer>, in HTML 5 provide a clear semantic meaning for document processing for humans and machines [6]. HTML is getting critical in the mobile-first digital environment, and responsive HTML technology provides an enhanced user experience in the mobile environment as well as the typical web environment [7]. HTML has been used to present content both on the web and on mobile using hybrid web technology.

Electronic documents based on HTML with responsive web technology offer an improved user experience and a suitable layout for various devices and interactive functions. However, electronic documents based on HTML are not single files and rely on related external resources. This can prevent downloading electronic documents based on HTML and store an image or PDF file instead. Additionally, HTML does not have content integrity verification specifications. It means there is no standard specification to distribute electronic documents based on HTML for the online and offline environment and no standard protocol to verify the content integrity of electronic documents based on HTML.

2.3. Content Integrity

When digital content is delivered from the creator to the recipient, it may be necessary for the recipient to verify that the delivered content is identical to the original. A one-way hash function is used to verify the content integrity. A one-way hash function is a function that generates a unique fixed-length message that is generally shorter than the original message for any size message. This one-way hash function is called a hash function or message digest function, and the fixed-length short hash value generated through this function is called a hash value, message digest, or fingerprint value [8]. This one-way hash function is ideal for verifying the integrity of the original message, as it generates a vastly different message digest even if the input message is slightly different. In addition, the generated message digest cannot be converted back to the original message, and it is impossible to reconstruct the original message. The integrity verification procedure of digital content using this characteristic is illustrated in Figure 1. The sender sends both the digital content and the hash value, which is generated through the hash function to the receiver. The receiver also generates the hash value for the received digital content using the same hash function and then compares the generated hash value with the received hash value to confirm if the digital content is original.

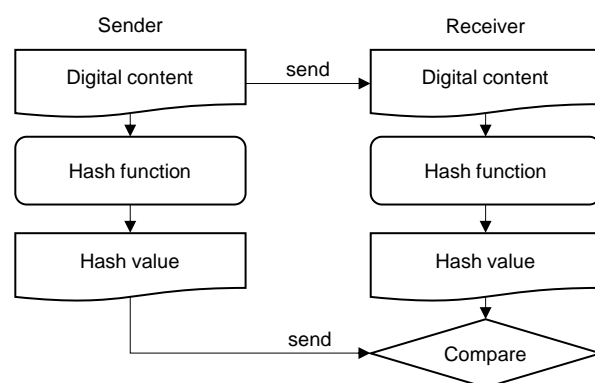


Figure 1. The procedure for verifying content integrity.

HTML presents content combining external resources in general, and it could be difficult to obtain a message digest including external resources. The message digest from

only the main HTML cannot guarantee the content integrity of full content. It is required to obtain a message digest of full content whether all the HTML resources are internal resources as a single file or calculate the message digest separately for each resource.

2.4. Digital Signature

A digital signature is a way of authenticating the identity of the sender of a digital message or document. A digital signature uses cryptographic key technology, which is a pair key structure with a public key and a private key. If someone wants to send an encrypted message to a receiver, they will use the receiver's public key to encrypt the message. Then, only the receiver will be able to decrypt and read the message. Similarly, when a sender wants to digitally sign a document, the sender uses the sender's private key to create a digital signature, and anyone with the sender's public key can verify that the signature was indeed created by the sender [9]. The RSA algorithm is used to create a public key and a private key. It defines a large prime number p and q , as shown in Figure 2, and calculates $N = pq$. Then, the public key $\{N, e\}$ is obtained by selecting a disjoint integer e after calculating $\phi(N)$. The private key $\{N, d\}$ is obtained by selecting an integer d after calculating $de = 1(\text{mod } \phi(N))$.

$p, q = \text{prime number}; p \neq q$
 $N = pq$
 $\phi(N) = (p - 1)(q - 1)$
 $\text{gcd}(\phi(N), e) = 1; 1 < e < \phi(N)$
 $de \text{ mod } \phi(N) = 1$
 $\text{public key} = \{N, e\}$
 $\text{private key} = \{N, d\}$

Figure 2. The structure of RSA algorithm.

This pair key structure with a public key and a private key is an asymmetric key structure, and the encrypted message using a private key can be decrypted using a public key. However, it is not possible to derive a private key from a public key. This means that a receiver can only decrypt an encrypted message and cannot encrypt messages. A sender keeps their private key and does not share the private key with anyone. A sender sends an encrypted message and the public key to a receiver, and a receiver decrypts an encrypted message using a public key, as shown in Figure 3.

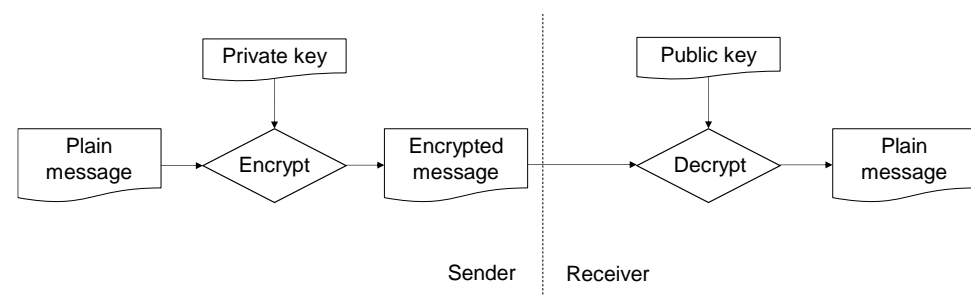


Figure 3. The encrypted message delivery.

A receiver needs to know the private key to encrypt a message. However, it took around four hours to compute an RSA-512-bit public key to obtain the private key using factoring in 2015. It predicted that it will take more than 4.006×10^{12} years to compute an RSA-2014-bit key [10]. Therefore, it is virtually impossible to find a private key for someone who only knows a public key for message encryption purposes. A receiver needs to secure the reliability of the keys when the encrypted message arrives. The decrypted message cannot have reliability without the reliability of the public key and private key. Therefore, the public key infrastructure guide states that all keys should be created by a

CA (Certificate Authority). A CA is a trusted third-party organization that verifies the certificate holder's identity and issues all the keys. A CA is linked to a Root CA, which is the top-level entity in the public key infrastructure. Therefore, all the keys from all the CAs have the same level of reliability as a Root CA. This means that a receiver reviewing the key creator information can have the reliability of the keys, and the message can be secured from that.

2.5. Document HTML and Chained Document

Electronic documents based on HTML have a vulnerability in terms of the content integrity perspective compared to a document based on PDF. There is no standard specification to verify content authenticity, and HTML is easily altered in an unauthorized manner. Document HTML is the proposed HTML specification to provide content integrity to act as a trusted electronic document [1]. Electronic documents should be a single independent file for convenient distribution, and this file has to have content integrity verification features. However, it is difficult to provide an electronic document based on HTML as a single independent file because HTML and resource files exist separately. Moreover, PDF can be distributed as a single independent file with content integrity verification features using a digital signature. Document HTML proposes restricted HTML specification, as shown in Table 2, to solve the weakness of an electronic document based on HTML that could not provide reliability. Chained Document is the proposed HTML specification based on Document HTML for electronic documents that must have a content integrity verification feature, and the document metadata is inserted into the Chained Document blockchain network to have reliable document distribution records.

Table 2. Conformance of Document HTML.

Conformance	Description
Document encoding	UTF-8
Resources	All resources must be embedded to be a single file, using Data URL Scheme in RFC 2397.
Multimedia tags	Multimedia tags, such as <audio> and <video>, are not allowed as these are not essential in terms of a document perspective, and they could cause a file size problem.
External resources container tags	<iframe>, <object>, <embed>, and <param> are not allowed as they bring content from an external location, and they could cause vulnerability in terms of content integrity.
Asynchronous data loading	Asynchronous data loading using script action is not allowed as it can change the content, and it could cause vulnerability in terms of content integrity.
Digital signature	The content must be digitally signed using PKI certificate.

However, the proposed Document HTML and Chained Document use the HTML comment declaration to contain the extended tag, as shown in Figure 4. This extended meta tag based on HTML comments is difficult to use in browsers because browsers ignore HTML comment tags while rendering HTML.

Moreover, a third-party document delivery service provider sends an electronic document to a receiver and provides a document delivery tracking mechanism [11]. A Certified Electronic Document Intermediary in Korea is authorized by the government, and they provide an electronic document distribution certificate to prove their document delivery service. Therefore, we have improved the Document HTML specification to focus on having a reliable electronic document with a third-party document delivery service provider.

```

<!--BEGIN DOCUMENT-HTML-TYPE
Document HTML or Chained Document declaration
END DOCUMENT-HTML-TYPE->
<!--BEGIN DOCUMENT-HTML-CONTENT-DIGEST
Hash digest of content
END DOCUMENT-HTML-CONTENT-DIGEST->
<!--BEGIN DOCUMENT-HTML-CONTENT-SIGNED-DIGEST
Digital Signed hash digest of content
END DOCUMENT-HTML-CONTENT-SIGNED-DIGEST->
<!--BEGIN DOCUMENT-HTML-CONTENT-VALIDATION-KEY
Digital certificate public key and information if format is Document HTML
END DOCUMENT-HTML-CONTENT-VALIDATION-KEY->
<!--BEGIN TRANSACTION-ADDRESS
Chained Document blockchain transaction address if format is Chained Document
END TRANSACTION-ADDRESS->

```

Figure 4. Extended tags of Document HTML and Chained Document.

3. Enhanced Document HTML

3.1. Improvement of Document HTML

The previously proposed Document HTML and Chained Document have vulnerabilities in terms of usability and content integrity, but we have improved the design of Document HTML. First, Document HTML and Chained Document contain the extended tag, which has limited usability of the digital signature as an HTML comment tag specification is used for the extended tags, as shown in Figure 4. Therefore, the extended meta tag is difficult to use as a browser does not render an HTML comment tag. The Enhanced Document HTML meta tags are defined by an HTML <meta> tag, as shown in Figure 5, so it has better usability in a web browser as it is an extended keyword definition in a <meta> tag. The targeted content to be digitally signed is located before and after the Enhanced Document HTML meta tag as meta tags are located in <head> tags, as shown in Figure 6. The <ds-range> meta tag has the byte position value for the target area, as shown in Table 3. It has four subsequent hexadecimal expressions, which are the start offset and end offset of the before-signature area and the start offset and end offset of the after-signature area.

```

<meta name="ds-range" content="byte position"/>
<meta name="ds-digest" content="labeled message digest"/>
<meta name="ds-signed-digest" content="signed message digest"/>
<meta name="ds-cert" content="certificate"/>

```

Figure 5. Enhanced Document HTML meta tags.

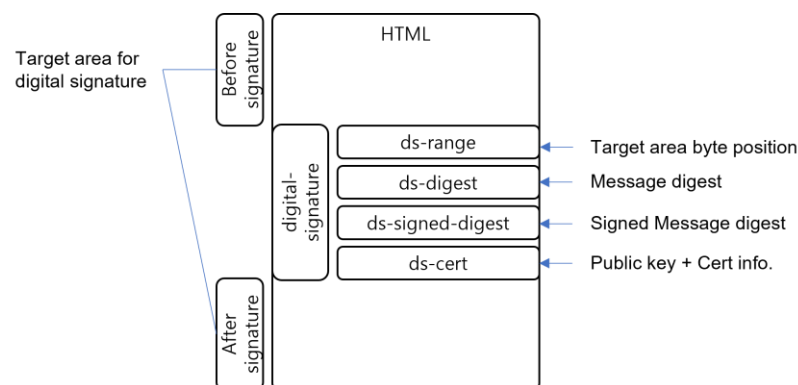


Figure 6. Enhanced Document HTML structure with the digital signature.

Table 3. <ds-range> meta tag in the Document HTML.

tag	"ds-range"
content	"Byte Position"
content format	"[0-9a-f]{8}[0-9a-f]{8}[0-9a-f]{8}[0-9a-f]{8}"
example	<meta name="ds-range" content="00000000000000FF00000AFF00000BFF"/>

Second, an Enhanced Document HTML type declaration is made by the standard DOCTYPE, contrary to the previous research that proposed a Document HTML type declaration using an HTML comment tag. It provides more usability and makes it easier for web browsers to understand the document type. Third, the @import function is not allowed in CSS. The @import function can link to an external CSS, and it is against the definition that all resources for Enhanced Document HTML must be embedded. The content of Enhanced Document HTML can be displayed differently if an external CSS is revised in an unauthorized manner. Therefore, it is not allowed.

Last, Enhanced Document HTML does not allow the use of script. The previously proposed Document HTML only does not allow the use of asynchronous data loading to avoid dynamic content loading. However, the non-asynchronous script function can also load the data from an external source, and it causes content integrity vulnerability. Therefore, it is not allowed.

3.2. Definition of Extended Document HTML

Extended Document HTML is the proposed HTML specification to provide more robust content integrity to act as a trusted electronic document. Extended Document HTML proposed a restricted HTML specification to solve the weakness of an electronic document based on HTML, which could not provide reliability through the following conformance based on improvement:

- (a) Extended Document HTML must have a DOCTYPE declaration, as shown in Figure 7;
- (b) Extended Document HTML uses UTF-8 encoding;
- (c) All resources must be embedded, and external resources are not allowed. The data URL Scheme in RFC 2397 is used to convert resources to internal resources, as shown in Figure 8;
- (d) The @import function is not allowed in CSS. The @import function can link to an external CSS, and it causes vulnerability in terms of content integrity. Therefore, it is not allowed;
- (e) Action script, such as JavaScript, is not allowed. The script function can load the data from an external source, and it causes vulnerability in terms of content integrity. Therefore, it is not allowed;
- (f) A multimedia tag, such as <audio> or <video>, is not allowed. A multimedia tag is not essential to present content in a document. These tags are not essential in terms of a document perspective, and they could cause a file size problem. Therefore, they are not allowed;
- (g) The <iframe> tag is not allowed. An <iframe> tag links to content from an external location, and the content is not part of the document. As such, it causes vulnerability in terms of content integrity. Therefore, it is not allowed;
- (h) An external resources container, such as <object>, <embed>, or <param>, is not allowed. These tags allow links to non-HTML objects. These tags contain device or OS-dependent values, making it difficult to embed them. Therefore, they are not allowed;
- (i) An Extended Document HTML meta tag must be included to have a content integrity verification feature, as shown in Figure 5. The <ds-range> tag indicates the byte area in the Document HTML, which is needed to have content integrity. The <ds-digest> tag has the message digest value using a hash function for the area. The <ds-signed-

digest> tag has the signed message digest value from the <ds-digest> value using a PKI certificate. The <ds-cert> tag has the public key and the certificate information to verify the <ds-signed-digest> value.

```
<!DOCTYPE document-html>
```

Figure 7. DOCTYPE declaration of Document HTML.

```
data:[<mediatype>][;base64],<data>
```

Figure 8. Data URL Scheme.

An Extended Document HTML specification makes an electronic document based on HTML a single independent document with a content integrity verification feature. It provides content integrity like a PDF does as it uses a PKI certificate. Extended Document HTML can have a responsive content presentation by inheritance from HTML and CSS technology and document authenticity using a PKI certificate. It means Extended Document HTML has an advantage over electronic documents based on both HTML and PDF. In addition, it is a suitable file format for the long-term archive as Extended Document HTML is an independent document format with embedded resources. However, there is no protocol or system to generate Extended Document HTML and verify Extended Document HTML in a legacy system. Therefore, an Extended Document HTML system could be required for generating and verifying Extended Document HTML, which integrates with a legacy system to deliver a reliable electronic document to a customer.

4. Design of Electronic Document Distribution Service Based on Enhanced Document HTML

4.1. Certified Electronic Document Intermediary

There are two major processes for distributing an electronic document in an enterprise. First, the electronic document content needs to be generated, called document generation. Second, the electronic document content needs to be delivered to a customer from an enterprise, called document delivery. Document delivery needs to be delivered via various channels, such as email, and a document delivery service provider does this. A document delivery service provider has to provide a secure platform to protect sensitive personal information. Therefore, a government or a central consortium organization manages the qualification of being a document delivery service provider [11]. The law is the “Framework act on electronic documents and transactions”, and there is a regulation for “Certified Electronic Document Intermediary” in the framework. There are fifteen certified electronic document intermediaries under the Korea Internet & Security Agency as of May 2022 [11]. These intermediaries provide delivery of digital content with user authentication, and they maintain the digital integrity metadata of the content. However, there is a vulnerability in content integrity as these services are unable to verify linked resources of the digital content based on HTML.

4.2. Electronic Document Distribution with Document HTML

Document HTML is a single electronic document format containing the metadata to verify content integrity. Therefore, it can secure content integrity with the document delivery service provider, the certified electronic document intermediary, as shown in Figure 9. The Electronic Document Creator requests to create Document HTML after creating an electronic document. Then, the electronic document is converted to Document HTML, which has a content integrity verification feature. The Electronic Document Creator stores Document HTML, and the link for the Document HTML is sent to the certified electronic document intermediary. The customer opens the Document HTML after the user

authentication by the certified electronic document intermediary, and the customer can see the content with content integrity verification features.

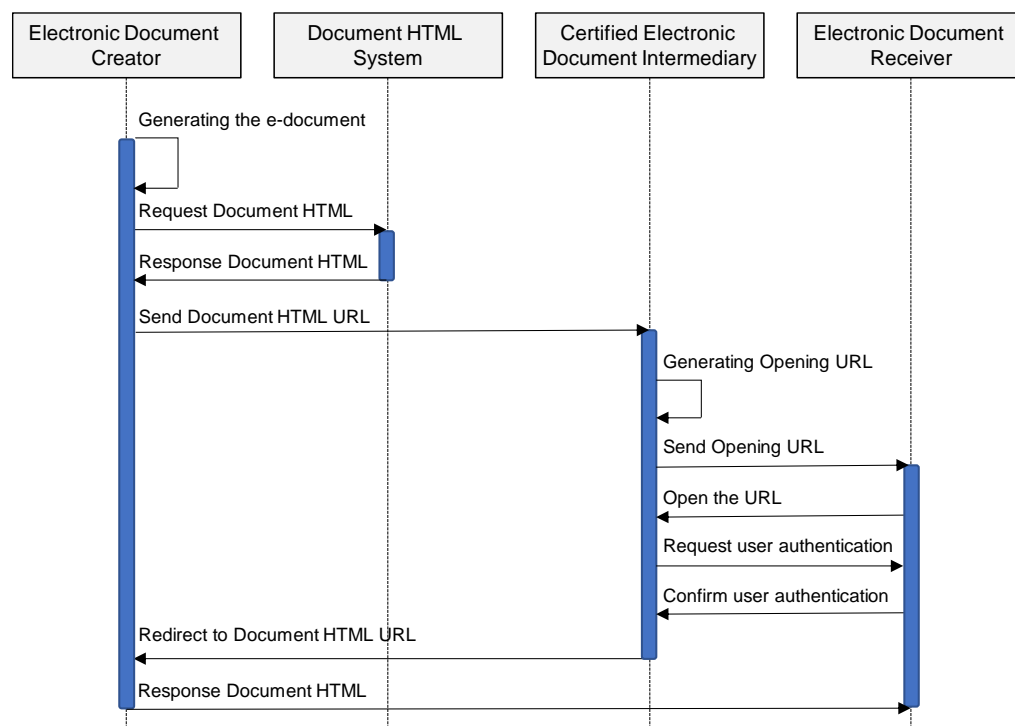


Figure 9. The E-Delivery with Document HTML system.

The integrated process with the certified electronic document intermediary and the Document HTML system provides the integrity advantage for the delivery and content perspective. A customer can secure the electronic document delivery verification by the certified electronic document intermediary and content integrity verification by the Document HTML system. Therefore, this integrated process can be used for the enterprise or government that must legally send the electronic document. It can be an alternative digital communication instead of registered postal mail.

5. Experimental Verification

5.1. HTML Electronic Document for Experiment

In our previous research, we used a dummy sample statement but, for this study, we have chosen the electronic tax notification document from the Korea National Tax Service. This electronic document is one of the highest volume documents sent to a citizen from the Korea Tax Agency and is delivered through a certified electronic document intermediary. Therefore, we have verified how HTML documents provide robust and reliable content integrity in real-world scenarios. The tax notification document can be opened after user authentication, and the certified electronic document intermediary redirects the URL link to the Korea National Tax Service website. Each URL link after user authentication is the personalized link that contains the user authentication metadata and is hard to predict to the URL for security. However, the tax notification document is based on HTML, linked to external resources, as shown in Table 4, and has a content integrity vulnerability. The tax notification document is opened in a web browser with linked external resources, as shown in Figure 10. There is no way to verify that the external resources are original.

The HTML tags in the tax notification document are shown in Table 5. All the tags are standard tags related to layout, such as the <p> tag, and there is no JavaScript action and <iframe> tag, which Document HTML does not allow the use of for content integrity purposes.

Table 4. Resources list for the personal tax notification document in Korea Tax Service.

URL	Type	Description
/jsonAction.do?actionId=action_id&tx_id=tx_id&token=tokenID	text/html	Main HTML
/img/rn/tmpl/10/2021/20220420_dt06.png	image/png	Common image
/img/rn/tmpl/10/2021/20220420_dt05.png	image/png	Common image
/img/rn/tmpl/10/2021/20220420_dt04.png	image/png	Common image
/img/rn/tmpl/10/2021/20220420_dt03.png	image/png	Common image
/img/rn/tmpl/10/2021/20220420_dt01.png	image/png	Common image
/img/rn/tmpl/10/2021/20220420_bot.jpg	image/png	Common image
/img/rn/tmpl/10/2021/20220420_at02.png	image/png	Common image
/img/cm/tmpl/img_tel.png	image/png	Common image
/img/cm/css/btn/btn_top.png	image/png	Common image
/img/cm/css/bg/ico_phone.png	image/png	Common image
/js/comm/jquery/jquery-ui.css	text/css	Style Sheet
/js/comm/iSwiper/swiper.min.css	text/css	Style Sheet
/css/comm/ntstb/styleTb.css	text/css	Style Sheet
/css/comm/ntstb/commonTb.css	text/css	Style Sheet
/css/comm/NtsCommonTb.css	text/css	Style Sheet
/css/comm/cm_style.css	text/css	Style Sheet

성실납세하시는 귀하께서 대한민국의 영웅입니다.

2021년 귀속 종합소득세 확정신고 안내

5월 HomeTax

D유형 (간편장부대상자)

▶ 신고기간 : 2022.5.1. ~ 5.31.
▶ 납부기간 : 2022.5.1. ~ 8.31.

코로나19 극복 지원을 위해 2022.5.31.까지 신고하는 경우 2022.8.31.까지 납부기한을 연장합니다.

님 안녕하십니까?

귀하의 성실납세에 감사드리며, 2021년 귀속 종합소득세 확정신고와 관련한 안내사항을 알려드립니다.

국세청에서는 홈택스(www.hometax.go.kr)를 통해 「종합소득세 신고도움서비스」를 제공해 드리니, 신고 전에 꼭 확인하여 주시기 바랍니다.

귀하의 성실납세에 감사드리며, 2021년 귀속 종합소득세 확정신고와 관련한 안내사항을 알려드립니다.

국세청에서는 홈택스(www.hometax.go.kr)를 통해 「종합소득세 신고도움서비스」를 제공해 드리니, 신고 전에 꼭 확인하여 주시기 바랍니다.

Content of the tax notification electronic document

DevTools is now available in Korean! Always match C

Elements Console Sources Netwo

Filter ☐ Invert ☐ Hide data U

☐ Has blocked cookies ☐ Blocked Requests ☐ 3rd-pa

☐ Use large request rows

☒ Show overview

10 ms 20 ms 30 ms 40 ms

Name	Status	Type
%EA%B5%AD%EC%84%B8%...	200	document
cm_style.css	200	stylesheet
20220420_dt01.png	200	png
20220420_at02.png	200	png
img_tel.png	200	png
20220420_dt03.png	200	png
20220420_dt04.png	200	png
20220420_dt05.png	200	png
20220420_dt06.png	200	png
20220420_bot.jpg	200	jpeg
ico_phone.png	200	png
btn_top.png	200	png

Loaded external resources

Figure 10. The Tax Notification Electronic Document.

Table 5. HTML Tags in the Tax Notification Document.

Tag	Count	Tag	Count	Tag	Count	Tag	Count
p	81	col	20	caption	8	title	1
td	79	br	19	colgroup	7	link	1
tr	61	strong	15	dt	4	html	1
div	58	ul	10	dd	4	head	1
th	46	tbody	8	thead	3	em	1
li	37	table	8	meta	3	button	1
span	21	img	8	dl	3	body	1

5.2. Generation of Document HTML

We generated the Document HTML using the sample tax notification document, and the generation result is shown in Table 6. Document HTML is generated well, as no items are being violated. The eleven external resources are converted into internal resources, and the tax notification document is generated into a single Document HTML. The file size has been increased from 1,833,435 bytes to 2,462,018 bytes because of the Document HTML metadata and BASE64 encoding of the internal resources. However, the file size is almost the same as the tax notification document with a PNG image file format for personal archiving purposes.

Table 6. The Document HTML Generation Result.

Item	Result
The converted external resources to the internal resources	11 resources (1 style sheet and 10 images)
File size of the original tax notification document	1,833,435 bytes
File size of the tax notification document as an image format for download purposes	2,414,571 bytes
File size of the Document HTML	2,462,018 bytes

The Document HTML has the same content as the original tax notification electronic document, and all resources are internal, as shown in Figure 11. Moreover, the Document HTML meta tags are inserted to verify the content integrity, as shown in Figure 12. A customer, who doubts the document's originality, verifies the content integrity in the verification menu from the Document HTML system using these Document HTML metadata.

5.3. Verification of Document HTML

We verified the Tax Notification Electronic Document, which was generated in the previous section and verified the harmed Tax Notification Electronic Documents to compare the verification result. We received the verification result via the verification web menu, as shown in Figure 13. Figure 13a shows that the Tax Notification Electronic Document is valid and keeps content integrity using the digital certificate issued by Let's Encrypt. Figure 13b shows the verification failures due to the content being altered after generating the document in Document HTML format. Additionally, Figure 13c shows the verification failures due to the digital signature being altered.

Thus, the Tax Notification Electronic Document based on Document HTML can be verified when a document receiver has to confirm the integrity of the document as Document HTML is a single HTML file with digitally signed extended meta tags.

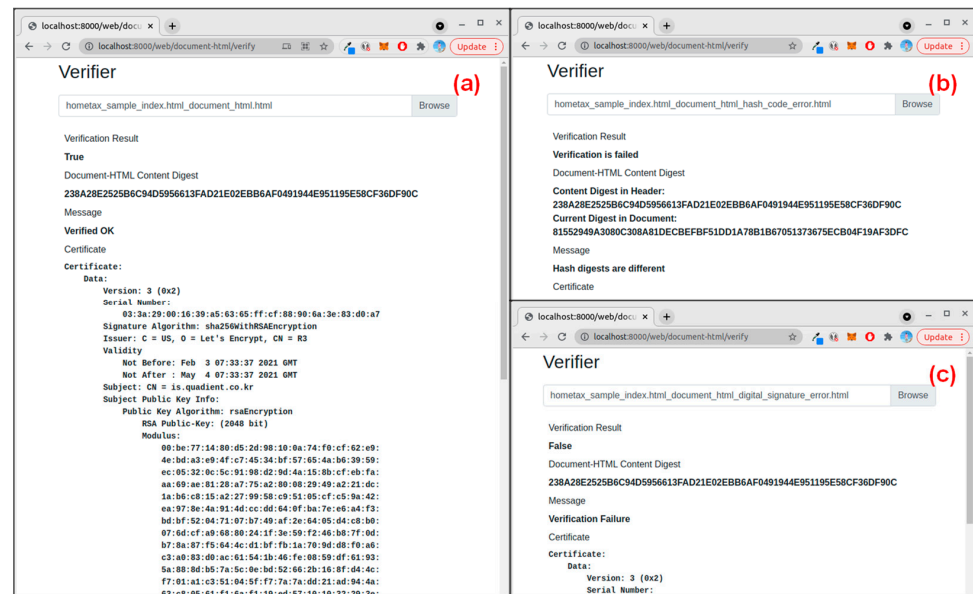


Figure 13. Verification example of the Tax Notification Electronic Document. (a) Verification result of valid document; (b) Verification result of invalid document; (c) Verification result of invalid digital signature.

6. Discussion and Limitations

Enhanced Document HTML is a type of digitally signed version of HTML. As there is no way to verify the content integrity of external resources in HTML, we proposed to embed all related resources internally and set conformance to remove the vulnerability content integrity perspective. In addition, Enhanced Document HTML can have responsive web content presentation by inheritance from HTML and CSS technology. It means Enhanced Document HTML has a flexible content layout for various devices, including mobile devices, and document authenticity. Enhanced Document HTML needs to work with third-party solutions as a service-oriented service because it helps to generate a reliable electronic document [12], as shown below in Figure 14. Most enterprise solutions are unified via the data integration layer, and each solution can generate a reliable electronic document with Enhanced Document HTML via the present integration layer [13].

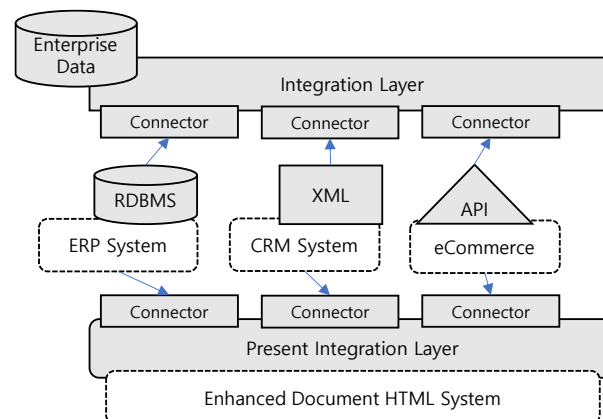


Figure 14. Data and present integration layer: basic architecture.

However, Enhanced Document HTML has strong definitions to provide a reliable electronic document. Hence, it has a limitation in providing an abundant interactive user experience. For example, it is not allowed to conduct dynamic content loading from a server in Enhanced Document HTML. Therefore, Enhanced Document HTML is helpful in

a particular domain that needs to send electronic documents containing static content with content integrity.

7. Conclusions

All enterprises and governments provide customer service. An electronic document is an essential communication tool for delivering content to a customer instead of physical documents in the digital era. Communication based on physical channels is well-established, and regulation has been developed to maintain content and distribution integrity. PDF is the de facto standard electronic document format for replacing a physical document, and its specification has been improved to provide content integrity. However, electronic documents based on HTML, the prevalent language in web and mobile environments, do not have standard content integrity verification features, so there is a vulnerability. We enhanced Document HTML to create a reliable single electronic document with a content integrity verification feature, and it has better usability in web browsers. We researched the generation and distribution of electronic document generation with a certified electronic document intermediary, and we designed an electronic document distribution service using both a certified electronic document intermediary and Document HTML for a real-world scenario. Additionally, we researched the tax notification document from the Korean National Tax Service and conducted experimental verification using the tax notification document. We confirmed electronic documents based on Document HTML are usable with third-party electronic document delivery service providers and provided a content integrity verification feature so that a customer can be sure an electronic document based on Document HTML has content integrity. We expect Document HTML to be used by enterprises and governments to deliver a reliable electronic document with a legal right to avoid legal disputes. In future research, we will continue to design service-oriented architecture to be one of the solutions in enterprise systems to provide reliable electronic documents.

Author Contributions: Writing—original draft, H.-C.H.; Writing—review & editing, W.-J.K.; Supervision, W.-J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Seoul National University of Science and Technology grant number 2020-0643.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hwang, H.C.; Kim, W.J. Design and Implementation of Document-HTML System for an Authorized Electronic Document Communication. *J. Adv. Eng. Technol.* **2021**, *14*, 61–73.
2. Hwang, H.C.; Kim, W.J. Design of Chained Document HTML Generation Technique Based on Blockchain for Trusted Document Communication. *Electronics* **2022**, *11*, 1006. [CrossRef]
3. Warnock, J.E.; Geschke, C. Founding and Growing Adobe Systems, Inc. *IEEE Ann. Hist. Comput.* **2019**, *41*, 24–34. [CrossRef]
4. PDF Association. PDF Specification Index. 2022. Available online: <https://bit.ly/3bpyZV1> (accessed on 25 May 2022).
5. HTML5. 2022. Available online: <https://html.spec.whatwg.org/> (accessed on 25 May 2022).
6. Lee, B.-K. HTML specification and semantics analysis of Korean news sites. *J. Digit. Contents Soc.* **2017**, *18*, 949–959.
7. Kaczmarczyk, A.; Zabierowski, W. The Comparison of Native and Hybrid Mobile Applications for Android System. In Proceedings of the 2021 28th International Conference on Mixed Design of Integrated Circuits and System, IEEE, Lodz, Poland, 24–26 June 2021; pp. 290–293.
8. Long, S. A Comparative Analysis of the Application of Hashing Encryption Algorithms for MD5, SHA-1, and SHA-512. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2019; Volume 1314, p. 012210.
9. Jun-Ho, S.; Sung-Su, K.; Seog, J.M. Diffie-Hellman Based Asymmetric Key Exchange Method Using Collision of Exponential Subgroups. Korea Information Processing Society. *Softw. Data Eng.* **2020**, *9*, 39–44.
10. Dasso, A.; Funes, A.; Riesco, D.; Montejano, G. Computing Power, Key Length and Cryptanalysis. An Unending Battle? *arXiv* **2020**, arXiv:2011.00985.
11. E-Document Integration Support Center. Certified Electronic Document Intermediary. Available online: <https://bit.ly/3y1shw8> (accessed on 25 May 2022).

12. Górski, T. UML Profile for Messaging Patterns in Service-Oriented Architecture, Microservices, and Internet of Things. *Appl. Sci.* **2022**, *12*, 12790. [[CrossRef](#)]
13. Petrasch, R.J.; Petrasch, R.R. Data Integration and Interoperability: Towards a Model-Driven and Pattern-Oriented Approach. *Modelling* **2022**, *3*, 105–126. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.