International Journal for Modern Trends in Science and Technology
Volume: 02, Issue No: 12, December 2016
ISSN: 2455-3778
http://www.ijmtst.com

Data Mining on Web URL Using Base 64 Encoding to Generate Secure URN

Nadia Ali[1] | M.Sharmila Devi[2]

[1]PG Scholar, Department of CSE, Geethanjali College of Engineering & Technology, Kurnool, Andhra Pradesh.
[2]Assistant Professor, Department of CSE, Geethanjali College of Engineering & Technology, Kurnool, Andhra Pradesh.

To Cite this Article
Nadia Ali, M.Sharmila Devi, "Data Mining on Web URL Using Base 64 Encoding to Generate Secure URN ", *International Journal for Modern Trends in Science and Technology*, Vol. 02, Issue 12, 2016, pp. 33-35.

ABSTRACT

The current Web has no general mechanisms to make digital artifacts such as datasets, code, texts, and images verifiable and permanent. For digital artifacts that are supposed to be immutable, there is moreover no commonly accepted method to enforce this immutability. These shortcomings have a serious negative impact on the ability to reproduce the results of processes that rely on Web resources, which in turn heavily impacts areas such as science where reproducibility is important. To solve this problem, we propose trusty URIs containing cryptographic hash values. We show how trusty URIs can be used for the verification of digital artifacts, in a manner that is independent of the serialization format in the case of structured data files such as nano publications. We demonstrate how the contents of these files become immutable, including dependencies to external digital artifacts and thereby extending the range of verifiability to the entire reference tree. Our approach sticks to the core principles of the Web, namely openness and decentralized architecture, and is fully compatible with existing standards and protocols.

KEYWORDS: *Data Mining, File Content Access, RDF Access, RDF Transferral.*

I. INTRODUCTION

Now a day's technology growth is huge specifically in digital technology, recreation is critical. Provable, unique, and originality are a vital fixing for making the results of mechanized procedures replicable, be that as it may, the present Web offers no ordinarily acknowledged strategies to guarantee these properties. Tries for example, the Web to distribute information in a digitized way shows the issue, in which digital calculations working on huge measures of information can be relied upon to be much more original than people to be controlled or manipulated substance. Without suitable counter-measures, unidentified attackers can harm or trap such calculations by including only a few deliberately controlled things to extensive arrangements of data information. To take care of this issue, we propose a way to deal with make things on the (Semantic) Web certain, unique, what's more, original. This methodology for Uniform Resource Identifiers (URIs) contains cryptographic hash values and holds fast to the standards of the Web; in particular openness furthermore, decentralized design. Proposed system is an implementation and feature work of paper [1]. This methodology for Uniform Resource Identifiers (URIs) contains encrypted notations and sticks to the standards of the Web, in particular transparent and fragmented design. Present paper we developed and updated form of a technical paper An encrypted notations(once in a while called cryptographic review) are short arbitrary

having succession of bytes (or, bits) which are ascertained way from an advanced artifacts[2], for example, a document. The same information dependably prompts the very same hash esteem, while only a negligibly altered data gives back a totally diverse quality. While there is an endlessness of conceivable inputs that prompt a particular given hash esteem, it is unthinkable practically speaking to remake any of the conceivable inputs just from the hash esteem. Present approach makes a difference to a particular and permanent advanced artifact.

II. Related Work

Existing system

In many areas and in particular in science, reproducibility is important. Verifiable, immutable, and permanent digital artifacts is an important ingredient for making the results of automated processes reproducible, but the current Web offers no commonly accepted methods to ensure these properties. Endeavors such as the Semantic Web to publish complex knowledge in a machine-interpretable manner aggravate this problem, as automated algorithms operating on large amounts of data can be expected to be even more vulnerable than humans to manipulated or corrupted content. Without appropriate counter-measures, malicious actors can sabotage or trick such algorithms by adding just a few carefully manipulated items to large sets of input data.

Disadvantages of Existing System:

1. Web content corrupted by human beings.
2. In existing, no methods to make web content immutable.

Proposed system

We propose an approach to make items on the (Semantic) Web verifiable, immutable, and permanent. This approach includes cryptographic hash values in Uniform Resource Identifiers (URIs) to the core principles of the Web, namely openness and decentralized architecture. Our proposed approach boils down to the idea that references can be made completely unambiguous and verifiable if they contain a hash value of the referenced
digital artifact. Our method does not apply to all URIs, of course, but only to those that are meant to represent a specific and immutable digital artifact.

Advantages of Proposed System:

1. We can make content on verifiable, immutable and permanent.

III. Implementation

Distributed Communication

Here each domain is free to communicate with each other, since the index of reference tree is built in multiple domain. A document can cite other publisher in a distributed environment. Each domain can post their publication to another domain or even to themselves. The domain to which posted can either approve or reject the document. But it requires no validation if posted to themselves. If approved its RDF [10] is automatically generated, and the document is published on the interface, accessible to all others across the network and if rejected its corresponding entry will be deleted. A domain himself acting as an attacker can sabotage the entrusted document given upon trust. But even presented in a distributed environment enabling inter-domain communication, the system ensures security to the document making digital artifacts on the web verified and trustworthy using trusty URIs and prevents MIMA attacks.

File Content Access

At FA, using SHA-512 hash generation algorithm [8] hash value is calculated, to which after appending two zero-bits are converted to Base64 notation generating trusty URN and complete trusty URI.

RDF Access

At RA, supports multiple graphs which works on RDF content. It allows self-references, resources that contain their own trusty URI. For Unicode characters a SHA-512 is generated in UTF-8 encoding, append two zero bits and is finally converted to Base64 notation

RDF Transferral

At RB, trusty URI represents single RDF graph. Similar to RA, hash value is calculated for Unicode using SHA-512 in UTF-8 encoding and is transformed to Base64 notation.

IV. Experimental Work

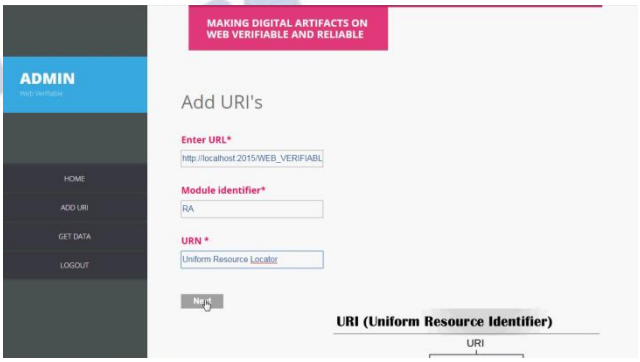

Fig:-1 Adding URL

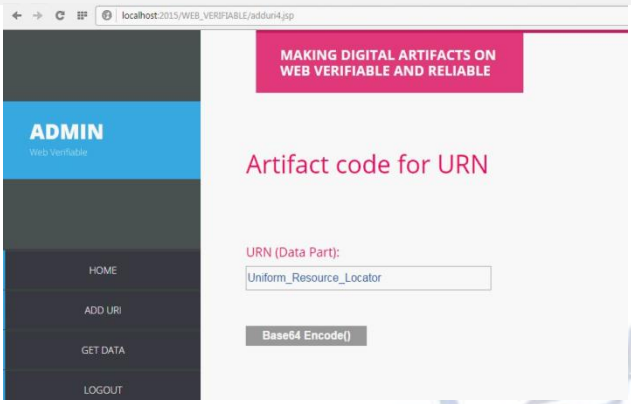

Fig: - 2 Base 64 Encoding on URN

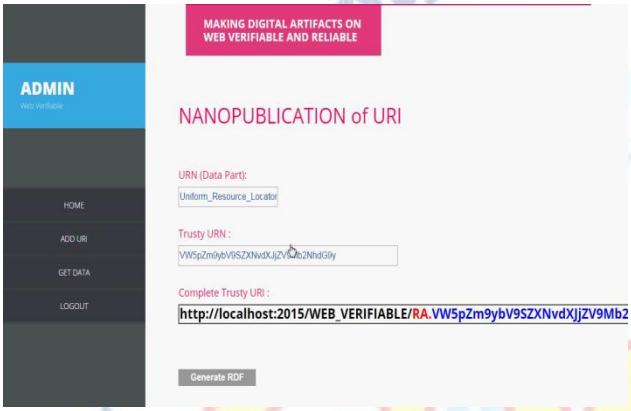

Fig: - 3 RDF Generator

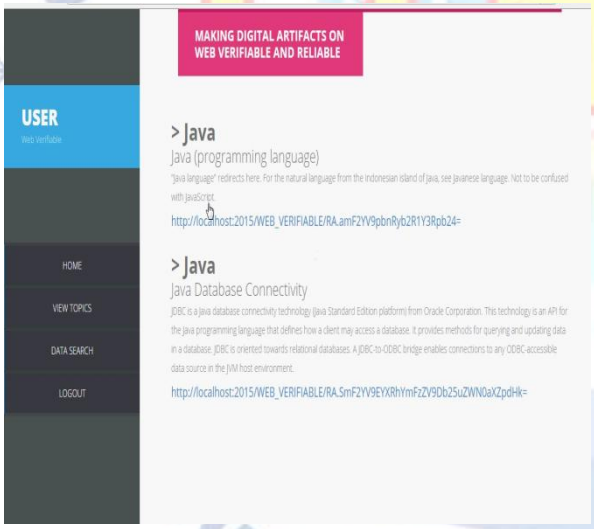

Fig: - 3 User Side Search

V. CONCLUSION

The Distributed digital artifact system for MIMA is where resources are distributed among different domains and each domain can communicate with each other. Unlike centralized system, distributed system gives users possibility to check whether the data have been modified. The relevant man in the middle attack is prevented to an extent by ensuring verifiability and reliability. The system ensures that data published within the system interface cannot be accessed anywhere outside the system, with the use of reference trees providing security at an overall level. Any manipulation to the data is efficiently identified and any further access to that data is blocked by informing user that the uniform location has been changed. Here only man in the middle attack is considered. This can be extended to more attacks.

REFERENCES

[1] Tobias Kuhn and Michel Dumontier, "Making Digital Artifacts on the Web Verifiable and Reliable", IEEE Transactions on Knowledge and Data Engineering, Vol NO 99 YEAR 2015

[2] Patrick P. Tsang, Apu Kapadia, Member, IEEE, Cory Cornelius, and Sean W. Smith, "Nymble: Blocking Misbehaving Users in Anonymizing Networks", IEEE Transactions ON Dependable and Secure Computing

[3] Tobias Kuhn and Michel Dumontier, "Trusty URIs: Verifiable, Immutable, and Permanent Digital Artifacts for Linked Data", in Proceedings of the 11th Extended Semantic Web Conference (ESWC 2014), ser.Lecture Notes in Computer Science. Springer, 2014

[4] N K Prasanna Anjaneyulu anna, Shaik Nazeer, "Semantic Web Security and Privacy", Journal of Theoretical and Applied Information Technology

[5] Tobias Kuhn, Christine Chichester, Michael Krauthammer and Michel Dumontier, "Publishing without Publishers:a Decentralized Approach to Dissemination,Retrieval, and Archiving of Data", arXiv preprint arXiv:1411.2749, 2014

[6] Momi Maity, Neha Verma, Rupali Wadikar, Sayali Shevkar, Prof. V.K. Bhusari, "Providing Security to Web Applications in Anonymizing Networks Using Nymble" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 1, January 2014 ISSN: 2277 128X

[7] Mingwu Zhang, Xiangyu Zhan, Sunil Prabhakar, "Cost Effective Forward Tracing Data Lineage", Computer Science Technical Reports. Paper 1669, 2007

[8] S.FarrelL, C.Dannewitz, D.Kutscher, B.Ohlman, A.Keranen, P. Hallam-Baker, "Naming Things with Hashes", Internet Engineering Task Force (IETF),April 2013

[9] Robert Ikeda and Jennifer Widom, "Data Lineage: A Survey", frmikeda,widomg@cs.stanford.edu

[10] C. Sayers and A. Karp, "Computing the digest of an RDF graph", Mobile and Media Systems Laboratory, HP Laboratories, Palo Alto, USA, Tech. Rep. HPL-2003-235(R.1),