Making Digital Artifacts On The Web Verifiable and Reliable By Using Cryptographic Hash Key

Naseema Shaik[1], Noha Abdullah ayedalshahrani[2], Afnan saadalali[2], Amjad mohammedsaad Alqahtani[2], Salhasaeedhedan[2]

[1] Lecturer, King Khalid University, KSA
[2] Student, King Khalid University, KSA

Abstract: *Currently we used single server to making digital artifacts and we are increasing burden to the server. To making digital resources on the web verifiable, immutable, as well as permanent, we have several techniques to comprise cryptographic hash values in URIs such as called as trusty URIs. At present the problems which we are facing are, there is no reliable standard procedure of verifying whether a received file truly represents the correct as well as original state of that artefact and by using single server to making digital artefacts and we are increasing burden to the server. In this paper we exhibit how the substance of these records get to be one of a kind. Our implementation holds good for the all levels in the Web, for example, transparent and fragmented engineering, which is completely good for present conventional models.*

Keywords: *Sematic Web, Nanopublications, Trusty URI's, Digital Technology.*

I. INTRODUCTION

Research Objects are an approach related to nanopublications, directing to establish ''self-contained units of knowledge'', and they constitute in a sense the antipode approach to nanopublications. We can call them mega-publications, as they contain much more than a typical narrative publication, namely resources like input and output data, workflow definitions, log files, and presentation slides[3]. In this paper we discuss, however, that bundling all resources of scientific studies in large packages is not a necessity to ensure the availability of the involved resources and their robust interlinking, but we can achieve that also with cryptographic identifiers and a decentralized architecture.

In this paper we include, most importantly, a new evaluation on the retrieval of nanopublication datasets over an unreliable connection, a description of the new feature of surface patterns, the specific protocol applied by existing servers, a server network that is now three times as large as before, a much more detailed walk-through example, and five new figures[1]. We furthermore present more details and discussions on topics including applications in the humanities, traversal-based querying, underspecified assertions, caching between architectural layers, and access of the server network via a web interface.

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 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 make a difference to a particular and permanent advanced artifact [4].

Here 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 [7]. 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.

II. BACKGROUND WORK

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[5]. 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[6]. Without appropriate counter-measures, malicious actors can sabotage or trick such algorithms by adding just a few carefully manipulated items to large sets of inputdata.

The concept of nanopublications has also been taken up in the humanities, namely in philosophy, and history/archaeology. A humanities dataset of facts is arguably more interpretive than a scientific dataset; relying, as it does, on the scholarly interpretation of primary sources[5]. Because of this condition, ‘‘facts’’ in humanities datasets called as prosopographies have often been called ‘‘factoids’’, as they have to account for a degree of uncertainty. Nanopublications, with their support for granular context and provenance descriptions, offer a novel paradigm for publishing such factoids, by providing methods for representing metadata about responsibilities and by enabling discussions and revisions beyond any single humanities project.

A well-known solution to the problem of individual servers being unreliable is the application of a decentralized architecture where the data is replicated on multiple servers. A number of such approaches related to data sharing have been proposed; for example, in the form of distributed file systems based on cryptographic methods for data that are public or private. In contrast to the design principles of the Semantic Web, these approaches implement their own internet protocols and follow the hierarchical organization of file systems. Other approaches build upon the existing BitTorrent protocol and apply it to data publishing, and there is interesting work on repurposing the proof-of-work tasks of Bitcoin for data preservation[8]. There exist furthermore a number of approaches to applying peer-to-peer networks for RDF data, but they do not allow for the kind of permanent and provenance-aware publishing that we propose below. Moreover, only for the centralized and closed-world setting of database systems, approaches exist that allow for robust and granular references to subsets of dynamic datasets.

III. A SURVEY ON “THE ANATOMY OF A NANOPUBLICATION”

As the amount of scholarly communication increases, it is increasingly difficult for specific core scientific statements to be found, connected and curated. Additionally, the redundancy of these statements in multiple fora makes it difficult to determine attribution, quality, and provenance. To tackle these challenges, the Concept Web Alliance has promoted the notion of nanopublications (core scientific statements with associated context). In this document, we present a model of nanopublications along with a Named Graph/RDF serialization of the model. Importantly, the serialization is defined completely using already existing community developed technologies. Finally, we discuss the importance of aggregating nanopublications and the role that the Concept Wiki plays in facilitating it.

Here, we have proposed an initial nano-publication model, a format instantiation, and how the Concept Wiki can be used to facilate aggregation. The format is based on existing community produced ontologies and technologies[9]. The role of the CWAformat working group is to specify a minimal common format for nano-publications that enables their aggregation and the correct preservation of the associated provenance. The CWA working group aims not to develop new specifications but instead to identify existing technology and formats that can be used for aggregating nano-publications.

IV. SYSTEM DESIGN AND ARCHITECTURE

This approach includes cryptographic hash values in the web URI's, particularly acceptance and decentralized design. The constructed believable URIs contains encrypted notations. The server network can be seen as an unstructured peer-to-peer network, where each node can freely decide which other nodes to connect to and which nanopublications to replicate [10]. The URI pattern and the hash pattern of a server define the surface features of the nanopublications that this server cares about. We called them surface features, because they can be determined by only looking at the URI of a nanopublication. For example, the URI pattern ‘http://www.STUniversity.org/’ states that the given server is only interested in nanopublications whose URIs start with the given sequence of characters. Additionally, a server can declare a hash pattern like ‘ AA AB’ to state that it is only interested in nanopublications whose hash in the trusty URI start with one of the specified character sequences. As hashes are represented in Base64 notation, this particular hash pattern would let a server replicate about 0.05% of all nanopublications. Nanopublication servers are thereby given the opportunity to declare which subset of nanopublications they replicate, and need to connect only to those other servers whose subsets overlap. To decide on whether a nanopublication belongs to a specified subset or not, the server only has to apply string matching at two given starting points of the nanopublication URI which means the first position and position 43 from the end as the hashes of the current version of trusty URIs are 43 bytes long, which is computationally cheap[9].

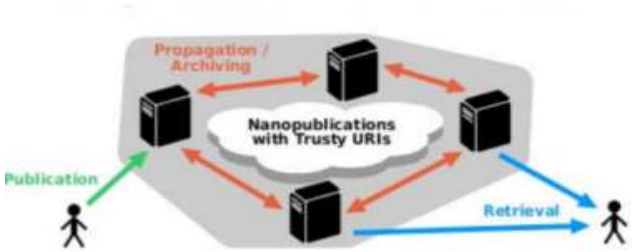

Fig.1 Architecture for Nanopublication Servers Network

Table 1 Comparison between the existing system and our application.

	Trusty URIs: Verifiable, immutable, and permanent digital artifacts for linked data	**The anatomy of a nanopublication**	**The MetaLex document server**	**Incremental cryptography: The case of hashing and signing**	**Digital Artifacts on Web with cryptographic control system**	**Making Digital Artifacts on the Web Verifiable and Reliable by using Cryptographic Hash Key**
Digital Artifacts are verifiable	×	✓	×	×	✓	✓
Data sets and Data Marts are immutable	×	×	✓	×	×	✓
Providing Security to protect data	×	×	×	✓	✓	✓
Using Hash Key Algorithm for secure browsing.	×	×	×	×	×	✓
Making the content permanent	✓	×	×	×	×	✓
URIs are trusty which are linked with digital artifacts.	×	×	×	×	×	✓

This is a structure of Nano publication servers. Such server's kind a server community, which can be used to post nano publications which have trusty URIs. One of these server best returns whole nano pubs; No queries supported; no triple retailer concerned. Nano publications are tiny snippets of data with provenance understanding attached. They may be able to be equipped in an extraordinarily flexible manner into colossal datasets which will also be described as nano publications.

V. CONCLUSION

We have presented a proposal for unambiguous URI references to make digital artifacts on the (Semantic) Web verifiable, immutable, and permanent. If adopted, it could have a considerable impact on the structure and functioning of the Web, could improve the efficiency and reliability of tools using Web resources, and could become an important technical pillar for the Semantic Web, in particular for scientific data, where provenance and verifiability are important. Scientific data analyses, for example, might be conducted in the future in a fully reproducible manner within "data projects" analogous to today's software projects. The dependencies in the form of datasets could be automatically fetched from the Web, similar to what Apache Maven (http://maven.apache.org) does for software projects, but decentralized and verifiable.

REFERENCES

[1] P. Groth, A. Gibson, and J. Velterop, "The anatomy of a nanopublication," Information Services and Use, vol. 30, no. 1, pp. 51–56, 2010.

[2] T. Kuhn and M. Dumontier, "Trusty URIs: Verifiable, immutable, and permanent digital artefacts for linked data," in Proceedings of the 11th Extended Semantic Web Conference (ESWC 2014), ser.Lecture Notes in Computer Science. Springer, 2014.

[3] R. Hoekstra, "The MetaLex document server," in The Semantic Web — ISWC 2011. Springer, 2011, pp. 128–143.

[4] M. Bartel, J. Boyer, B. Fox, B. LaMacchia, and E. Simon, "XML signature syntax and processing," W3C, Recommendation, June 2008. [Online]. Available: http://www.w3.org/TR/xmldsig-core.

[5] E. Hofig and I. Schieferdecker, "Hashing of rdf graphs and a ¨ solution to the blank node problem," in 10th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW 2014), 2014, p. 55.

[6] 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), 2004.

[7] R. Phan and D. Wagner, "Security considerations for incremental hash functions based on pair block chaining," Computers & Security, vol. 25, no. 2, pp. 131–136, 2006.

[8] J. McCusker, T. Lebo, C. Chang, D. McGuinness, and P. da Silva, "Parallel identities for managing open government data," IEEE Intelligent Systems, vol. 27, no. 3, p. 55, 2012.

[9] A. Callahan, J. Cruz-Toledo, and M. Dumontier, "Ontology-based querying with Bio2RDF's linked open data," Journal of Biomedical Semantics, vol. 4, no. Suppl 1, p. S1, 2013.

[10] J. Broekstra, A. Kampman, and F. Van Harmelen, "Sesame: A generic architecture for storing and querying RDF and RDF schema," in The Semantic Web — ISWC 2002. Springer, 2002,pp. 54–68.