Anatomy of a Semantic Virus

Peyman Nasirifard

Digital Enterprise Research Institute
National University of Ireland, Galway
IDA Business Park, Lower Dangan, Galway, Ireland
peyman.nasirifard@deri.org

Abstract. In this position paper, I discuss a piece of malicious automated software that can be used by an individual or a group of users for submitting valid random noisy RDF-based data based on predefined schemas/ontologies to Semantic search engines. The result will undermine the utility of semantic searches. I did not implement the whole virus, but checked its feasibility. The open question is whether nature inspired reasoning can address such problems which are more related to information quality aspects.

1 Introduction and Overview

Semantic-Web-Oriented fellows encourage other communities to generate/use/share RDF statements based on predefined schemas/ontologies etc. to ease the interoperability among applications by making the knowledge machine-processable. The emergence of semantic-based applications (e.g. Semantic digital libraries[1], SIOC-enabled shared workspaces[2], Semantic URL shorten tools [3]) and also APIs (e.g. Open Calais[4]) etc. are good evidences to prove the cooperation among application developers to talk using the famous subject-predicate-object notion. However talking with the same alphabets but various dialects brings ambiguity-related problems which have been addressed by some researchers and are out of scope of this paper.

Searching, indexing, querying and reasoning over (publicly) available RDF data bring motivating use cases for Semantic search engine fellows. The crawlers of Semantic search engines crawl the Web and index RDF statements (triples) they discover on the net for further reasoning and querying. Some of them are also open to crawl the deep Web by enabling users to submit the links to their RDF data.

Since the birth of computer software, especially operating systems, clever developers and engineers benefited from software security leaks and developed

[1] http://www.jeromedl.org/
[2] http://www.bscw.de/
[3] http://bit.ly/
[4] http://www.opencalais.com/

software viruses which in some cases brought lots of disasters to governments, businesses and individuals[5].

In this paper, I describe a potential piece of software which can be used by a malicious user or a group of synergic malicious users in order to undermine the utility the Semantic search engines. In brief, what the virus does, is generating automatically random noisy knowledge which will be indexed by Semantic search engines. My main motivation of presenting this idea here is identifying some research challenges in trust layer of the well-known Semantic Web tower.

It is worthwhile mentioning that the title of this paper *Anatomy of a Semantic Virus* is perhaps misleading. Actually, I am not going to describe the anatomy of a virus that is based on the Semantic Web[6], but rather I focus on a distributed virus that targets Semantic Web data.

The structure of this position paper proceeds like the following: In the next part, I describe the problem and a scenario that demonstrates the method that the potential virus may operate upon. In section 3, I have a discussion on potential directions of finding solutions. Finally, I conclude this short position paper.

2 Problem

Semantic Web search engines (e.g. SWSE[7], Swoogle[8]) crawl and index new Semantic Web documents containing RDF statements. There are some services available on the net (e.g. Ping The Semantic Web[9] (PTSW)) that enable end users to publicly submit and announce the availability of their Semantic Web data. These submissions can be later fetched by Semantic search engines for indexing and further reasoning.

The main module of the potential virus is a piece of code that receives as input several triples and generates as output several triples based on the inputs and also predefined schemas, so that the generated RDF triples are syntactically correct, but semantically wrong (fake). Figure 1 shows a simple example. As illustrated in the figure, the input is two RDF triples: "Galway is part of Ireland" and "London is part of England". The RDF schema has already defined that Galway and London are instances of the concept City, whereas Ireland and England are Countries. In this example, the virus exchanges the object (or subject) parts of triples, taking to account the fact that both objects (or subjects) are instances of the same class (Country or City). The generated result will be "Galway is part of England" and "London is part of Ireland"; which both are correct RDF statements, but wrong (fake) knowledge. Note that the whole process is done by a malicious software and it is not kind of supervised editing and/or does not have human in the loop.

[5] `http://www.landfield.com/isn/mail-archive/2000/May/0067.html`

[6] I see this a bit strange, as common computer viruses do not communicate to each other and interoperability among viruses is not well-defined.

[7] `http://swse.org/`

[8] `http://swoogle.umbc.edu/`

[9] `http://pingthesemanticweb.com/`

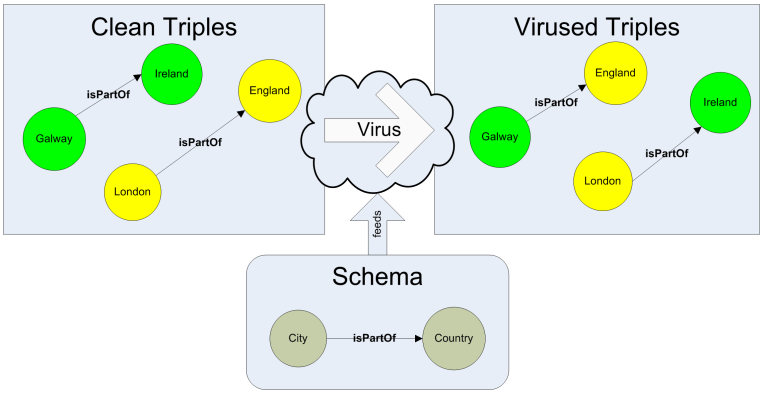

Fig. 1. Main module of the virus

The number of fake clones that can be generated is all possible instances of various concepts within RDF document.

Note that the same problem may exist on the Web and somebody may put fake knowledge on common Web pages. Moreover there exist lots of tricks to get high ranking in search engines, but as we know, the growth of the Semantic Web is not as fast as the Web[10] [1] and such malicious activities are feasible and can be *performed* using available RDF documents. Meanwhile, Semantic Web's main promise is to make the knowledge machine-processable, whereas the unstructured data on the Web is more suited for humans and obviously current machines do not have the wisdom and sense of humans.

On the other hand, someone may claim that due to the success of collaborative information gathering platforms like Wikipedia [11], the motivation for producing wrong knowledge in RDF is weak. However, we all benefit from platforms like Wikipedia, but we rarely use its articles to cite in scientific papers. The reason is perhaps the fact that the authors of such articles are unknown and we can not really trust on the content. The same applies to the RDF data. If we gather a large amount of Semantic Web data in RDF, can we really trust them? How to exclude potential fake triples from the knowledge base?

2.1 Scenario

Here I present a simple scenario to describe the possible attack that a virus can affect RDF data. As we know, publicly available services like PTSW, provide processable feeds that include the recently-added/updated RDF documents. These feeds are used by malicious software. However, the virus may even use Semantic search engines to find RDF data from the net.

[10] `http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html` and `http://sw.deri.org/2007/06/ontologymap/`

[11] `http://www.wikipedia.org/`

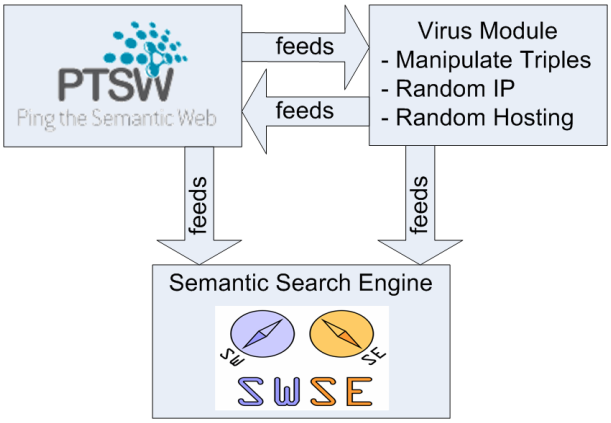

Fig. 2. Possible attack

In our scenario, the malicious software will parse the output feed of the PTSW and get an index of the published RDF files. Then it fetches the RDF statements from the net and changes them so that the generated RDF will be still valid. The result will be then submitted as an updated (or new) RDF after a random time interval with a random IP address (TCP/IP level) using a random hosting to PTSW which will be indexed by Semantic search engines. The malicious software may even submit the content directly to semantic search engines, if they provide such functionality. Figure 2 demonstrates the overall view of the possible attack which can be performed using PTSW service.

The main problem arises, when a group of people or even an individual in large scale employs several instances of the malicious software and generates fake RDF triples which will be submitted/indexed to/by the Semantic search crawlers. If search engines are not capable to cope with this situation, the result will undermine the utility of semantic searches.

3 Discussion

Digital signature is a vertical layer in the Semantic Web tower. There exist some third-parties that issue certificates for authorized users. However we may use digital signatures, certificates or any other means to cope with authentication and authorization aspects of RDF data, but we can not cope with the Quality aspects of the information (accuracy, validity, etc.). Moreover, we can not really bound the usage of Semantic Web to only authenticated, authorized and/or certified parties. Otherwise, we are highly eliminating its usage.

It is important to consider that the source of a piece of data is an important factor in validity and accuracy which are two important concepts of information quality. However, the virus is not able to change the origin of RDF document, but it is able to edit the RDF with fake statements. As *virused* RDF is still

valid based on a schema, it can not be simply tracked for possible manipulation. One research problem that arises with this issue is investigation on the cloned graphs to find out the original one and perhaps log the cloned versions as illegal graphs. Probably one naive approach is using a trusted knowledge body (universal common sense facts) that verify the material generated by others. But maybe this also brings some limitations and we do not have a really comprehensive knowledge base for the whole universe facts. On the other hand, nature inspired reasoning tries to benefit from other domains to address mainly the complex reasoning challenges within Semantic Web. The open question is whether nature inspired reasoning can be useful in this area to validate the quality aspects of data.

To my view, this problem and its potential solutions can bring also some commercial interests. As an example, building a trusted knowledge party that can validate RDF-based knowledge generated by people or giving authorities to people to evaluate (semi-automatically) the generated RDF-based knowledge by others.

4 Conclusion

In this position paper, I presented briefly a method that can be used by a piece of automated software to maliciously target Semantic Web data, in order to put lots of noisy elements into the knowledge base. I mentioned that the search results of Semantic search engines may not be really trustable, as they may contain fake noisy knowledge and machines can not really benefit from them, unless we are certain that the existing knowledge in their repositories is true reliable one.

The fact that I presented this idea here is exploring some research problems that I am not aware of their solutions, after reviewing literature and having some discussions with senior Semantic Web researchers. Generating meaningful clones of a given graph based on a schema (virus) and identifying the original one from a bunch of cloned graphs (anti-virus) are possible research directions that can be further explored. I personally did not implement the whole virus, but I checked its feasibility using PTSW and a set of fake triples.

Acknowledgments. I thank Vassilios Peristeras and Stefan Decker for their supports. This work is supported by Ecospace (Integrated Project on eProfessional Collaboration Space) project: FP6-IST-5-35208

References

1. Ding, L., Finin, T.: Characterizing the Semantic Web on the Web. Proceedings of the 5th International Semantic Web Conference, 2006.