Modeling Vulnerable Internet of Things on SHODAN and CENSYS : An Ontology for Cyber Security

Marc Arnaert
Univ. Nice Sophia Antipolis, I3S, UMR 7271, 06900 Sophia Antipolis, France
CNRS, I3S, UMR 7271,
06900 Sophia Antipolis, France
e-mail: marc@arnaert.com

Yoann Bertrand
Univ. Nice Sophia Antipolis, I3S, UMR 7271, 06900 Sophia Antipolis, France
CNRS, I3S, UMR 7271,
06900 Sophia Antipolis, France
e-mail: bertrand@i3s.unice.fr

Karima Boudaoud
Univ. Nice Sophia Antipolis, I3S, UMR 7271, 06900 Sophia Antipolis, France
CNRS, I3S, UMR 7271,
06900 Sophia Antipolis, France
e-mail: karima@unice.fr

***Abstract*—With the increase of connected devices on the Internet, managing security can become a very difficult task for Information Technology (IT) and security managers. In order to find vulnerabilities for these devices, we can use search engines. Despite the fact that these engines are very powerful, they often propose various and complex syntaxes for queries. Moreover, sorting the results, due to their complexity and quantity, can be challenging and time-consuming for IT & security managers. To overcome these issues, we propose in this paper an ontology that can reduce the complexity and improve the results of search engines to help these managers to detect vulnerable devices.**

Keywords-vulnerability; Shodan; Internet of Things; Censys; ontology; cyber security

I. INTRODUCTION

US research firm Gartner has estimated that 20.8 billions objects will be connected on the Internet by 2020 [1]. This tremendous amount of objects, often referred as Internet of Things (IoT), can be problematic for security.

Indeed, in order to be accessible and operational, these objects need to be connected (for instance, video surveillance, telephony, building management systems, air conditioners, automated doors, etc.). Such quantity of devices can leave a potential open door that can be exploited by intruders.
Moreover, IoT components are heterogeneous and often poorly protected, increasing *de facto* the risk of security breaches (for instance, Fiat-Chrysler has recalled 1.4 Million cars to prevent hacks in July 2015) [2].

To secure connected objects within their companies, security experts or employees in charge of defining security policies can gather information in order to build a vulnerability assessment plan and have a better understanding of the weaknesses of the deployed devices.

In other words, an expert wants to answer the following question: "*Is my device vulnerable and accessible by someone else?*".

To do so, she/he can use specialized search engines available on the Internet. However, due to the increasing quantity, known vulnerabilities and diversity of connected objects, a basic search can return thousands of results. Such quantity of results can be complex to sort, understand and analyze, especially for non-security experts. Thus, this task can be difficult and time-consuming.

In this article, we propose an ontology that models vulnerabilities of IoT objects. Our ontology is based on existing search engines (Shodan and Censys) and aims at:

- Reducing the number of aggregated results during a search.
- Increasing the relevance of results (i.e., returned vulnerabilities).
- Be usable for non-security experts.

By doing so, we aim at reducing the time-consumption and increase the robustness and feasibility of assisted vulnerability assessments in IoT.

The rest of the paper is organized as follows: Section II presents the related works in the domain of security search engine. Section III describes our contribution, while Section IV discusses future works.

II. RELATED WORKS

A. Online Databases and tools

At first, the hackers used specialized security tools software to find potential and vulnerable target, like Kali-Linux [6] the most advanced penetration distribution.

In order to gather information for cybersecurity now, online databases and search engines can be used.

Shodan.io is a search engine designed by programmer John Matherly in 2009. It interrogates devices ports and grabs the resulting banners, then indexes the corresponding public IP address and search into an intern databases for futures lookup. Shodan aggregates a significant amount of information (more than 3.7 billion public IPv4 addresses and also checks hundreds of millions of IPv6 addresses).

 ISBN: 978-1-61208-493-0

Many wonderful works can be found on Shodan. These works encompass vulnerability assessment tool [4], reveals magnitude of IoT [5] and usages in industrial context [3].

Censys.io is a search engine designed by Zakir Durumeric in 2015 [7]. It allows researchers to ask questions about what composes the Internet. Censys collects data through daily Zmap scans of more than 3 Billions IP v4 addresses. Researchers can interact with these data through scripts that can be requested thanks to a SQL engine.

However, using Censys or Shodan can lead to the following issues:

- Results can be too numerous to be efficiently interpreted.
- Results can be irrelevant (i.e., outdated, non-specific, incomplete, etc.).
- Both queries and results can be hard to understand and analyze by non-security experts (i.e., you must know different syntaxes to interact with the two engines, you must sort the results to find the corresponding domain devices, you don't know if these devices are vulnerable and if yes, which vulnerabilities).

B. *Security Ontology*

Existing security ontology covers special domains like networks, architectures, cryptography or resilience domains. Few academic works have been proposed to use security ontology with Shodan ([8][9]) and no ontology has been actually published for Censys. None of them cover the semantic security knowledge to easily find vulnerable devices using Shodan or Censys.

Concerning the IoT domain, several specific ontologies have been proposed (i.e., Iot-ontology [10], SAREF [11] or openiot-ontology [12]). Again, none of them contain the semantic knowledge that could be used with Shodan or Censys to find vulnerable IoT devices into specific domain.

In conclusion, by using existing search engines, we need to know the semantic behind the search engine (in our case, Censys and Shodan) and need to be a security expert to find and understand vulnerabilities on these objects.

Moreover, existing search engines suffer from the following drawbacks:

- There is no correlation with CVE (Common Vulnerabilities Exposure) published by CERTs (Computer Emergency Response Team) [5].
- It is complex to achieve cross-domain and interdatabases operations.
- There is no possibility to reuse the results.
- There is no modeling or automated process.
- Errors or omissions can be done due to human interpretation.
- There is an excessive consumption of time to obtain exploitable results with vulnerable objects.

To overcome previous drawbacks, we propose an ontology that is described in the next section.

III. CONTRIBUTION

As stated previously, we first aim at reducing the quantity of results returned by Censys and Shodan. More specifically, we aim at reducing the complexity of semantic of returned results, increase the relevance of vulnerable objects and ease the use of such tools.

A. *Genericity of the proposal*

Although our contribution focuses on these two search-engines, our model has been implemented to be as generic as possible. Thus, our model can be adapted to the preferred search engine of the IT manager (i.e., Google, Bing, etc.). To facilitate this adaptation, we have used the concept of ontology.

B. *Definition of a cybersecurity ontology*

In order to propose a tool able to model this knowledge semantic, we have created an ontology of research and diagnosis. A graphical representation of such mechanism is depicted in Figure 1. This modeling will allow us to: reuse the field of knowledge, facilitate interoperability and portability, use a reasoning engine to bring new knowledge by inferences, add more accurate descriptions of metadata and facilitate its integration and update in an open and flexible way.

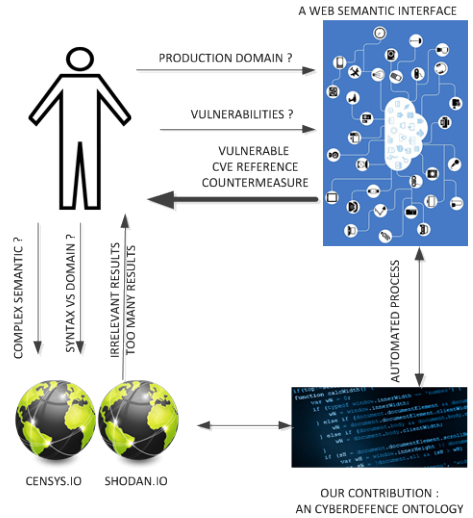

Figure 1. Specific and relevant results can be obtained via ontology.

To define our model, we have retained semantic web best practices and tools such NeOn methodology [13], Linked Data best practices [14].

 ISBN: 978-1-61208-493-0

These works propose methodologies to build well-structured ontologies or datasets from scratch and suggest reusing as much as possible existing works by linking them together.

We have also chosen Stanford's ontology editor 'Protégé 2000' to easily define and improve our ontology. This software is particularly recommended to create a domain ontology [15].

To conclude the main objectives of our work are:

- To create an ontology of research and diagnostic;
- To model a semantic search over Censys or Shodan;
- To include aggregation of existing security vulnerabilities ontologies;
- To validate the syntax of our ontology with W3C (World Wide Web Consortium) tools;
- To evaluate our ontology via standardization tools available on the Internet.

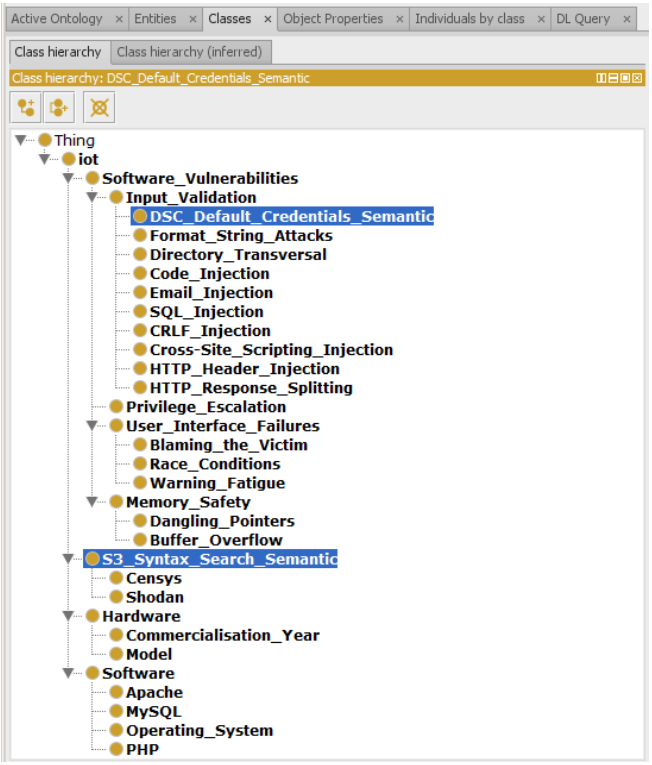

Figure 2. A first version of our ontology into 'protégé 2000'

a) Select semantic for specific domains

To find the semantic related to a specific domain, we have explored publications that define vulnerability semantic [16], projects focusing on Shodan [3], darknet videos [17], deposits of sites (Pastebin), website of published vulnerabilities (db-exploit), and other GoogleDork [18].

The aggregation of these information allowed us to create a primary knowledge base called **S^3 "Syntax Search Semantic"** who can target specific objects in relation to specific research areas. Effectively, using the keywords stored in S^3 in engines such as Censys and Shodan allows users to find the right targeted devices.

b) Select reference vulnerability

In order to test our ontology, we have chosen a reference vulnerability type. For our preliminary tests, we have chosen the "*default credentials*" vulnerability. This vulnerability exploits default logins and passwords. These credentials can easily be found in the manufacturers' documentation that are accessible on the Internet and contain default passwords in clear texts.

Moreover, we have find default credentials lists on the Internet (1000+) and have aggregated them to increase the number of possibilities into our database.
To facilitate our search, we have developed a script that searches documents for specific keywords (i.e., "username", "password", "login", etc.). Thanks to this script, we have generated a database of existing manufacturer's default credentials.

We assume that the person who installed the device has not changed the default password, and we can use it successfully. The advantage of this vulnerability is that it is not iterative and therefore not intrusive and somewhat wordy, in the sense that we do not try repeatedly to penetrate a system. We have created a second database that keeps the "*default credentials*" of associated devices. We will call this second database DCS "Default Credentials Semantic ".

Figure 2, shows a first version of our ontology where S^3 and DCS knowledge can be highlighted to help us to find a field type of vulnerable and assailable IoT device.

c) Validation of S^3 with DCS

We have tested the correlation between S^3 and DCS over Censys. Our ontology will be available on Internet with a specific namespace (URI). It contains S^3 and DCS information. The realization of a first program in Python (both allow on Shodan & Censys), allows us to test our ontology for which the test proves that it is fully functional.

We have obtained at screen a short list of public IP addresses with the open type protocol. Figure 3. shows a textual example of the obtained results. For these preliminary tests, we have validated our proposal by doing a copy/paste of one of the public IP addresses returned into a new browser page. By entering this IP, we have reached the login page of the connected IoT device.

 ISBN: 978-1-61208-493-0

```
CENSYS Request Vulnerable Objects with a research and diagnostic Cyberdefence Ontology
Scientific contribution - Marc ARNAERT -> Tuesday 15th of March 2016 01:07:17 PM

Semantic Search Syntax in Censys.io :
---------------------------------------------------------------------
[1] IP: 211.76.13      - Protocol : [u'443/https', u'21/ftp']
[2] IP: 211.21.17       Protocol : [u'80/http', u'995/pop3s', u'25/smtp', u'110/pop3',
[3] IP: 211.21.17       Protocol : [u'80/http', u'443/https', u'21/ftp']
[4] IP: 180.43.28       Protocol : [u'80/http', u'21/ftp']
[5] IP: 202.213.4       Protocol : [u'80/http', u'21/ftp']
[6] IP: 219.103.9       Protocol : [u'80/http', u'21/ftp']
[7] IP: 81.227.90      Protocol : [u'80/http', u'110/pop3', u'21/ftp', u'443/https', u'
[8] IP: 60.249.18       Protocol : [u'80/http']
[9] IP: 106.187.9      - Protocol : [u'80/http', u'22/ssh', u'53/dns']
[10] IP: 210.59.1      - Protocol : [u'80/http']
[11] IP: 210.242.      - Protocol : [u'80/http', u'110/pop3', u'443/https', u'25/smtp',
[12] IP: 59.120.1       - Protocol : [u'80/http', u'110/pop3', u'21/ftp', u'443/https',
[13] IP: 118.99.2      - Protocol : [u'80/http', u'993/imaps', u'995/pop3s', u'25/smtp'
[14] IP: 60.248.1       - Protocol : [u'80/http', u'53/dns']
---------------------------------------------------------------------
Number of Vulnerable IOT found : 14
```

Figure 3. Finding Vulnerable Objects in Censys with our application

For these preliminary tests, we have validated our proposal by doing a copy/paste of one of the public IP addresses returned into a new browser page. By entering this IP, we have reached the login page of the connected IoT device.

We have fetched the corresponding login and password in our DCS for this specific device. By entering these credentials, we have obtained a granted access to the full configuration of the IoT device, proving that our solution is quite functional.

In final, we have built a Sparql server [20] and validate basic Sparql queries to obtain the complete list of S^3 and DCS.

IV. FUTURE WORK

In this paper, we have proposed an ontology to obtain more accurate results concerning vulnerable IoT devices, when using search engines such as Censys or Shodan.

For future works, we will create a user-friendly semantic web application that has two goals. The vulnerable domain researching will be unique and user-friendly (i.e., easy to understand and mastered for all types of users). Secondly, the type of vulnerabilities will aim at sorting and obtaining more relevant results.

ACKNOWLEDGMENT

We are extremely grateful to John Matherly for unlimited access to Shodan resources and his precious help and we thank Zakir Durumeric for his wonderful tool Censys.io.

We thank Olivier Corby and Isabelle Mirbel for their help concerning semantic web, Sparql syntaxes and ontologies.

We thank for their inducements, advices and valuable feedbacks Alain Giboin and Frédéric Precioso.

REFERENCES

[1] GARTNER, S. (2015). Gartner Says 6.4 Billion Connected “Things” will be in use in 2016, up 30 Percent from 2015 [interactive].[viewed 2015 m. november 10 d.]. *Access through internet: <http://www. gartner. com/newsroom/id/3165317*.

[2] Miller, C., & Valasek, C. (2015). Remote exploitation of an unaltered passenger vehicle. *Black Hat USA*.

[3] Bodenheim, R., Butts, J., Dunlap, S., & Mullins, B. (2014). Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices. *International Journal of Critical Infrastructure Protection*, *7*(2), 114-123.

[4] Genge, B., & Enăchescu, C. (2015). ShoVAT: Shodan-based vulnerability assessment tool for Internet-facing services. *Security and Communication Networks*.

[5] Radvanovsky, B. (2013). Project shine: 1,000,000 internet-connected scada and ics systems and counting. *Tofino Security*, *19*.

[6] Arnaert, M. (2015). Initiating to Ethical Hacking with Kali-Linux. Amazon press.

[7] Durumeric, Z., Wustrow, E., & Halderman, J. A. (2013, August). ZMap: Fast Internet-wide Scanning and Its Security Applications. In *Usenix Security* (Vol. 2013).

[8] Stepanova, T., Pechenkin, A., & Lavrova, D. (2015, September). Ontology-based big data approach to automated penetration testing of large-scale heterogeneous systems. In *Proceedings of the 8th International Conference on Security of Information and Networks* (pp. 142-149). ACM.

[9] Krotofil, M., & Gollmann, D. (2013, July). Industrial control systems security: What is happening?. In *Industrial Informatics (INDIN), 2013 11th IEEE International Conference on* (pp. 670-675). IEEE.

[10] Kotis, K., & Katasonov, A. (2013). Semantic interoperability on the internet of things: The semantic smart gateway framework. *International Journal of Distributed Systems and Technologies (IJDST)*, *4*(3), 47-69.

[11] Daniele, L., den Hartog, F., & Roes, J. (2015). Created in Close Interaction with the Industry: The Smart Appliances REFerence (SAREF) Ontology. In *Formal Ontologies Meet Industry* (pp. 100-112). Springer International Publishing.

[12] Soldatos, J., Kefalakis, N., Hauswirth, M., Serrano, M., Calbimonte, J. P., Riahi, M., ... & Skorin-Kapov, L. (2015). Openiot: Open source internet-of-things in the cloud. In *Interoperability and Open-Source Solutions for the Internet of Things* (pp. 13-25). Springer International Publishing.

[13] Suárez-Figueroa, M. C., Gomez-Perez, A., & Fernandez-Lopez, M. (2012). The NeOn methodology for ontology engineering. In *Ontology engineering in a networked world* (pp. 9-34). Springer Berlin Heidelberg.

[14] Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 205-227.

[15] Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. *University of Standford*.

[16] Leverett, E. P. (2011). Quantitatively assessing and visualising industrial system attack surfaces. *University of Cambridge, Darwin College*.

[17] Shovgenya, Y., Skopik, F., & Theuerkauf, K. (2015, June). On demand for situational awareness for preventing attacks on the smart grid. In *Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), 2015 International Conference on* (pp. 1-4). IEEE.

[18] Lancor, L., & Workman, R. (2007, March). Using Google hacking to enhance defense strategies. In *ACM SIGCSE Bulletin* (Vol. 39, No. 1, pp. 491-495). ACM.

[19] Cheng, J., Ma, Z. M., & Tong, Q. (2015). RDF Storage and Querying: A Literature Review. *Handbook of Research on Innovative Database Query Processing Techniques*, 460.

[20] Arnaert, M. (2016). Create your own SPARQL web Server for semantic web with DEBIAN & VIRTUOSO. Amazon press.

 ISBN: 978-1-61208-493-0