6 Security Architecture Techniques
6-1 Cybersecurity Information Discovery Technique and Knowledge Base

Takeshi TAKAHASHI

This paper proposes a scheme that links assorted structured cybersecurity information over networks and discovers them in order to enable sharing of cybersecurity information beyond the borders of organizations and countries. The scheme links assorted XML-based online repositories and provides a means to locate information across them. The scheme is unique in its flexible and extensible information structure. This paper also introduces its prototype implementation.

1 Introduction

In order to assure cybersecurity, information sharing that transcends organizational and national boundaries is required. To achieve this, several institutions are providing different types of cybersecurity information on the Internet. Typical examples of this are the National Vulnerability Database (NVD)[1] and Japan Vulnerability Notes (JVN)[2]. It is hoped that hereafter, organizations from all over the world will provide more information. However, at present there are problems with understanding where all this information is located, and it is difficult for individual users to extract and utilize only the information that is relevant to themselves from such a huge volume of information. To facilitate information sharing, discovery technologies are needed to enable users to know where such information is located and to find, identify, and obtain the necessary information.

This paper proposes discovery technology that can identify different types of cybersecurity information present on a network, and that can search and exchange that information. In the proposed method, the structure of the meta information used for search features flexibility and extendibility, which are achieved by defining information in two levels: category and format. As such categories, it uses the information categories[3] defined in the cybersecurity information ontology which was the author's previous study, while using the schema standardized by different organizations for formats.

This paper also introduces the implementation of the prototype of the proposed method. With this implementation, various repositories can be searched across the Internet. For more details, refer to the document by this author[4] that this paper summarizes.

2 Related research

A common format is needed for organizations to exchange/share cybersecurity information among themselves. Various organizations are already looking into techniques for describing the structuring of cybersecurity information. For example, CVE[5] defines the identifier and XML description technique of vulnerability information, and there are also ARF[6], CAPEC[7], CCE[8], CEE[9], CPE[10], CRF[11], CVRF[12], CVSS[13], CWE[14], CWSS[15], IODEF[16], MAEC[17], OCIL[18], OVAL[19], XCCDF[20], etc.

RDF[21] is a typical method used for identifying information on a network. RDF is a W3C standard for describing information on resources, and for identifying/searching resources on the Internet. RDF can be used to describe arbitrary entities. SPARQL[22] is defined for achieving search engine functions of RDF, and there are also various implementations.

This paper proposes a method for identifying and searching different types of security information, using the technology described above.

3 Design philosophy of the proposed method

The proposed method should structure the different types of cybersecurity information present on a network, making it possible to identify, search and exchange the information. The proposed method is built based on the following basic policy.

a) The search target is limited to information in XML format: There is already a variety of information described using XML, such as CVE or CAPEC, and the proposed method will handle only information provided in XML format. It provides free keywords search, and tag based search.

b) Maintain extendibility: Currently, there is a limited number of schemas of cybersecurity information in XML format, but the number is expected to increase in the future. Therefore, it is necessary to maintain extendibility that can also support future schemas.

c) Assure scalability: For now, there are limited amounts of information, but the amount is expected to increase in the future. Therefore, information needed should be searchable regardless of the amount of information.

d) Utilizing information in the existing XML format as is: A new protocol is required for searching information properly, but when considering implementation and deployment, it is important to have no requirement to make any modifications in the information itself which is currently provided online by the information source.

e) Security assurance technology of the proposed method itself is out of scope: There are various existing technologies such as encryption and authentication technology, which must be incorporated when the proposed method is put into practice, but these are not covered in this paper.

4 Architecture

As shown in Fig. 1, the proposed method consists of 4 roles: D-Client, D-Server, Registry and InfoSource. Each role is described in detail below. Note that there are also cases where one entity provides multiple roles.

- D-Client: This role communicates with D-Server and searches cybersecurity information. If required, it communicates with more than one D-Server.
- D-Server: This role is the interface for D-Client, and according to the request from D-Client, it searches for the appropriate URI of InfoSource. In the process, the D-Server communicates with one or more Registries and summarizes all their responses.
- Registry: This role is the interface for InfoSource; for registration requests from InfoSource, it collects/stores the metadata related to that InfoSource. Information in InfoSource can be identified based on this metadata.
- InfoSource: This role maintains the cybersecurity information described in the XML format. In order for this information to be searchable on a network, its own information is registered in one or more Registries.

5 Structure of information

This section specifies the data structure to be used in the architecture defined in Section **4**. The proposed method searches and discovers cybersecurity information. To accomplish this on a network, the information must be machine readable. Therefore, the proposed method uses various existing information formats. However, there are various information formats, they are expected to be changed in the future, and still further developments are foreseen.

To assure extendibility of these formats, the proposed method provides a technique wherein the data structure is divided and defined into two levels of information: category and format. In order to establish information categories that are highly abstracted and do not require changes for a long time, we use the information categories stipulated in the ontology of document [3]. Also, for formats to be actually used easily, explicitly defined schemas need to be used, so schemas defined by various industrial standards are used.

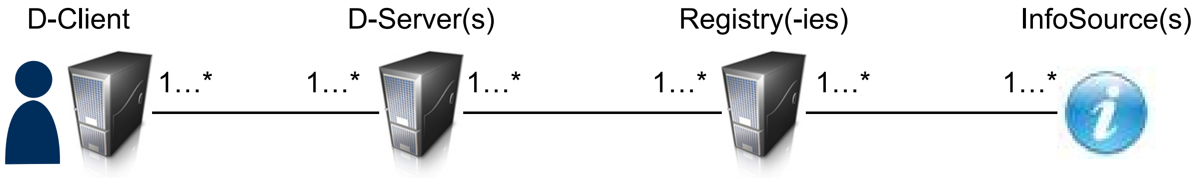


Fig. 1 Group of roles that constitute the proposed method

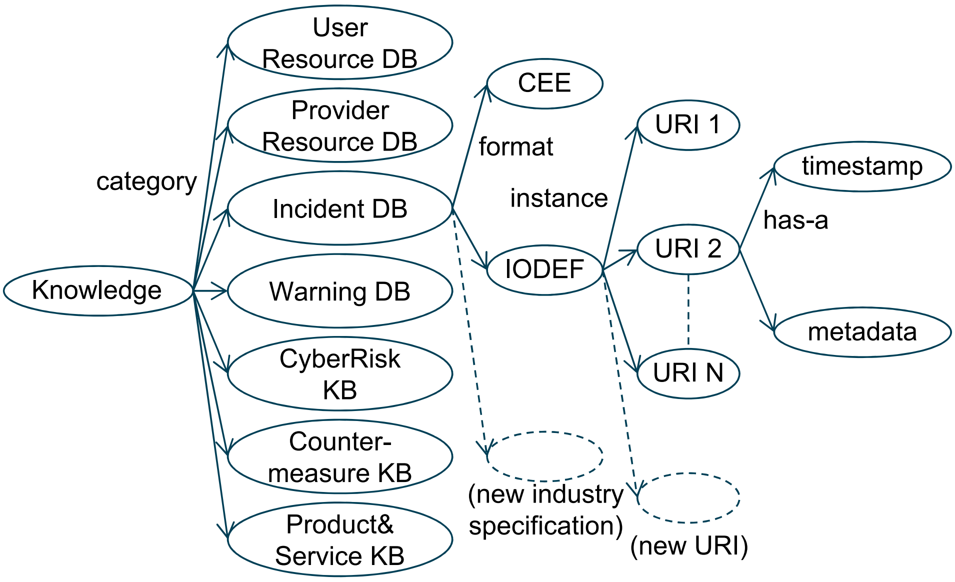


Fig. 2 Two-level data structure comprised of category and format

Figure 2 shows the data structure of the proposed method. All the information related to InfoSource is saved using URIs as the keys, those URIs are arranged according to the expression format, and those formats are arranged according to information category. Also, in each URI, at least the timestamp that indicates the date and time of information registration of InfoSource is linked to the metadata needed for search.

With this two-level structure and implementation in the RDF, the proposed method assures future extendibility. When using new schema in the future, it will suffice to just link the new schema with any of the above mentioned information categories, and it is easy to set up such links in the information structure managed by RDF.

6 Protocol

This section defines three types of procedures that the proposed method needs to find information: register information, register and cancel server, and search information.

6.1 Register information

This procedure is mandatory when InfoSource releases information. InfoSource sends the registration message for the Registry. This message contains the information related to URI of InfoSource and information category. When receiving this message, Registry accesses that URI and acquires the XML format information that is in that InfoSource. Registry executes XSLT for that information and generates and saves/updates the metadata of the information related to that InfoSource in RDF format.

When the metadata is saved/updated, the Registry sends the notification message to the D-Server. The updated information of the metadata (including the URI of that InfoSource) is included in this message. The D-Server can also send notification message to the registered D-Client, and thus the D-Client can acquire the updated security information faster.

6.2 Register and delete registration of server

The server registration procedure is needed when Registry selects D-Server. The Registry sends the join message to the D-Server to be used. The D-Server that received the message embeds the supporting information format and the information related to the information category which the information format belongs to in a result message and returns it. This paper proposes a single category that complies with the ontology of document [3], but it is also possible to arbitrarily specify the category and information format to be used to return this result message.

When the Registry wants to stop using the server, it must delete the server registration by either sending a leave message to the corresponding server, or waiting for the timeout period in D-Server. In either case, the corresponding server sends the result message to the Registry, to inform that it will suspend providing the service to the Registry.

D-Client can also implement this procedure, and it is possible to know beforehand the information category and format that the server supports.

6.3 Search information

In this procedure, D-Client acquires the URI of the InfoSource in which the necessary information is maintained, and defines the procedure for receiving information. InfoSource sends a query message to D-Server, and D-Server transfers that message to all the registered Registries. Each Registry receives the message, searches metadata related to the self-maintained InfoSource, ranks the InfoSource candidates, puts that ranking information into the result message, and returns it to the D-Server. The D-Server that received the result messages from all the Registries summarizes all the messages into a single message, puts that message into a single result message, and sends it to D-Client. D-Client selects one InfoSource from all InfoSource candidates described in the received result message, accesses the URI of that InfoSource, and acquires the desired information in XML format.

7 Prototype implementation

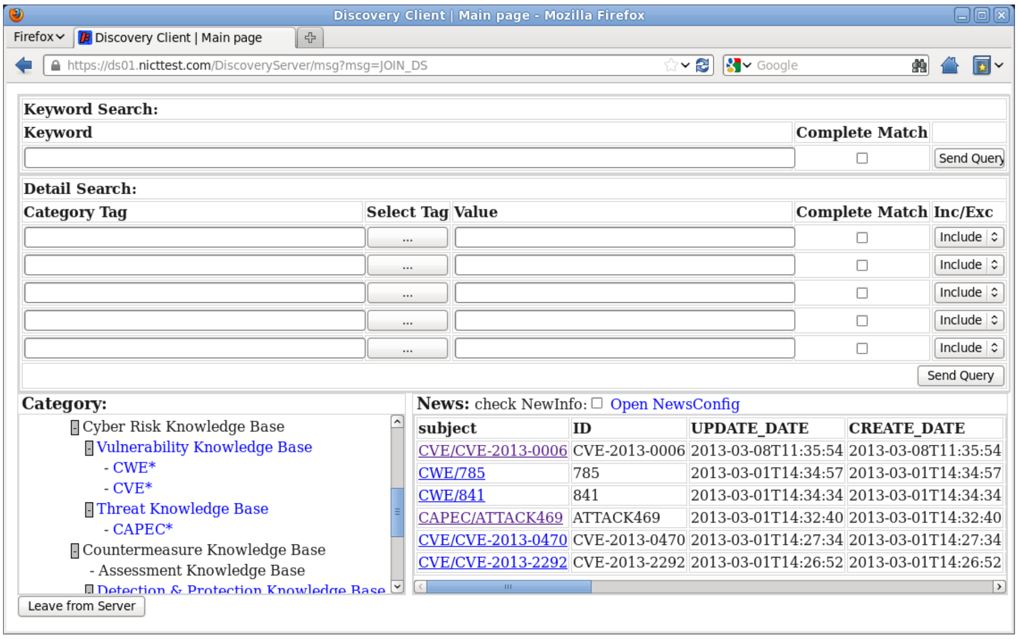
This section introduces implementation of the prototype of the proposed method. D-Client, D-Server, Registry and InfoSource are all implemented using Java, and run on CentOS. Further, Sesame[23], which is one of the implementations of the SPARQL engine, is used to search and manipulate RDF data. The following is an overview of this implementation.

Figure 3 shows web interface prepared for when D-Client searches using D-Server. Although this system can operate basically without this web interface because it is RESTful implementation, we prepared this interface to make the system more user friendly.

This screen provides four types of information collection means. First, it provides a free text search function. Using this function, the user can search by entering the character strings in any manner. As the second means, the screen provides a tag specific search function. Similar to the free text search, the user can search by entering the character strings in any manner, but it is also possible to narrow down the XML tags to be searched. Multiple combinations of search conditions are also possible. The third means is a category search function. Users can obtain a list of the desired information by specifying the category and then the format of the information they want to acquire. The last means is the latest information search function. This function displays a list of only the latest registered information. Note that, because of RESTful implementation as mentioned above, this system is available for use regardless of the web interface described above.

8 Observations

In this section, the extendibility and implementation aspects of the proposed method are observed. At the same time, it introduces the international standardization activities for making this technology available to society.



Discovery Client | Main page - Mozilla Firefox

https://ds01.nicttest.com/DiscoveryServer/msg?msg=JOIN_DS

Keyword Search:

Keyword | Complete Match □

Detail Search:

Category Tag | Select Tag | Value | Complete Match | Inc/Exc

Category:

- Cyber Risk Knowledge Base
  - Vulnerability Knowledge Base
    - CWE*
    - CVE*
  - Threat Knowledge Base
    - CAPEC*
- Countermeasure Knowledge Base
  - Assessment Knowledge Base
  - Detection & Protection Knowledge Base

News: check NewInfo: □ Open NewsConfig

subject	ID	UPDATE_DATE	CREATE_DATE
CVE/CVE-2013-0006	CVE-2013-0006	2013-03-08T11:35:54	2013-03-08T11:35:54
CWE/785	785	2013-03-01T14:34:57	2013-03-01T14:34:57
CWE/841	841	2013-03-01T14:34:34	2013-03-01T14:34:34
CAPEC/ATTACK469	ATTACK469	2013-03-01T14:32:40	2013-03-01T14:32:40
CVE/CVE-2013-0470	CVE-2013-0470	2013-03-01T14:27:34	2013-03-01T14:27:34
CVE/CVE-2013-2292	CVE-2013-2292	2013-03-01T14:26:52	2013-03-01T14:26:52

Fig. 3 Search screen using D-Server

8.1 Observations on extendibility and implementation aspects of the proposed method

The proposed method achieves a flexible structure in order to be compatible with various information schemas that will be registered in the future. When information structures based on existing standards become inconvenient and are less frequently used, it will suffice to just define and standardize a new information schema and link it with any of the information categories defined within the data structure of this method. Categories defined in ontology are used to avoid the need to change information categories in the near future, but if a category must be changed, another category can be used as mentioned above. Based on the above, it seems that the proposed method maintains extendibility.

Next, let us look into the points to consider when actually deploying the proposed method on the Internet. The content written in this paper has been designed only with the primary objective of exchanging cybersecurity information, and it does not cover safe usage of this technology. However, this technology handles cybersecurity information, which can be misused and cause great damage. Therefore, when implementing/using this technology, it is imperative that sufficient attention be given to security aspects. For example, countermeasures can be implemented using various existing techniques such as authenticating information registrants and users, encrypting communications, etc.

Also, for the prototype there is still an issue of its process speed. Among the four roles, Registry is the bottleneck in processing. It will be prudent if the Registry is implemented in a form that distributes the load as needed.

8.2 Standardization activities

The proposed method aims to implement information exchange between multiple organizations. To achieve this, a minimal common interface must be defined. One such interface is the various information schemas which were also introduced in previous research. For example, defining an information schema as an international standard can contribute to normalized, more efficient information exchange between organizations. The author is also actively involved in international standardization activities, and has contributed to the standardization of various technologies in ITU-T and IETF. In particular, ITU-T RecommendationX.1500[24] which establishes the framework of information exchange, IODEF-SCI[25] which establishes the schema extension of incident information, and ITU-T RecommendationX.1570[26] which establishes the framework of the discovery technique, are closely linked with the proposed method, and can be referred to.

9 Conclusion

The proposed method structures the different types of cybersecurity information present on a network, and identifies/searches/exchanges that information. The information structure of the proposed method features flexibility and extendibility, which are achieved by defining information in two levels: category and format. Also, implementation of the prototype has shown that the proposed method actually works. This method can facilitate the exchange of cybersecurity information across organizations/countries, and boost global cybersecurity. In the future, we hope to build a durable system for the actual operations. In this regard, countermeasures against security problems such as misuse of the system must be implemented, and they must be taken up as challenges for the future.

Acknowledgments

We wish to extend our heartfelt gratitude to Dr. Koji Nakao, Senior Researcher, Dr. Kazumasa Taira, Research Center Director, and Mr. Bhola Panta, Engineer, for their support in conducting this research.

References

1 National Institute of Standards and Technology, “National Vulnerability Database Version 2.2,” 2014. [Online]. Available: http://nvd.nist.gov/.

2 JPCERT/CC and IPA, “Japan Vulnerability Notes,” 2014. [Online]. Available: http://jvn.jp.

3 T. Takahashi, Y. Kadobayashi, “Reference Ontology for Cybersecurity Operational Information,” The Computer Journal, 2015.

4 T. Takahashi, Y. Kadobayashi, “Mechanism for Linking and Discovering Structured Cybersecurity Information over Networks,” IEEE International Conference on Semantic Computing, 2014.

5 International Telecommunications Union, “Common vulnerabilities and exposures,” ITU-T Recommendation X.1520, 2014.

6 National Institute of Standards and Technology, “Specification for the Asset Reporting Format 1.1,” NIST Interagency Report 7694, 2011.

7 International Telecommunications Union, “Common attack pattern enumeration and classification,” ITU-T Recommendation X.1544, 2013.

8 National Institute of Standards and Technology, “Common Configuration Enumeration (CCE),” [Online]. Available: http://nvd.nist.gov/cce/index.cfm. [Last access: 2014].

9 The MITRE Corporation, “Common Event Expression,” [Online]. Available: http://cee.mitre.org/. [Last access: Jan 2014].

10 International Telecommunications Union, “Common platform enumeration,” ITU-T Recommendation X.1528, 2012.

11 The MITRE Corporation, "Common Result Format Specification Version 0.3," [Online]. Available: http://crf.mitre.org/. [Last access: Jan 2014].

12 Industry Consortium For Advancement of Security on the Internet, "The Common Vulnerability Reporting Framework v1.1," [Online]. Available: http://www.icasi.org/cvrf-1.1. [Last access: Jan 2014].

13 International Telecommunications Union, "Common vulnerability scoring system," ITU-T Recommendation X.1521, 2011.

14 International Telecommunications Union, "Common weakness enumeration," ITU-T Recommendation X.1524, 2012.

15 International Telecommunications Union, "Common Weakness Scoring System," ITU-T Recommendation X.1525, 2015.

16 The Internet Engineering Task Force, "The Incident Object Description Exchange Format," RFC 5070, dec 2007.

17 International Telecommunications Union, "Malware attribute enumeration and characterization," ITU-T Recommendation X.1546, 2014.

18 National Institute of Standards and Technology, "Specification for the Open Checklist Interactive Language (OCIL) Version 2.0," NIST Interagency Report 7692, 2011.

19 International Telecommunications Union, "Language for the open definition of vulnerabilities and for the assessment of a system state," ITU-T Recommendation X.1526, 2014.

20 International Organization for Standardization/International Electrotechnical Commission, "Specification for the Extensible Configuration Checklist Description Format (XCCDF) Version 1.2," ISO/IEC 18180:2013, 2013.

21 The World Wide Web Consortium, "Resource Description Framework (RDF): Concepts and Abstract Syntax," http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/, 2004.

22 The World Wide Web Consortium, "SPARQL query language for RDF," http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/ .

23 openRDF.org, "SESAME," [Online]. Available: http://www.openrdf.org/. [Last access: March 2012].

24 International Telecommunications Union, "Overview of Cybersecurity information exchange (CYBEX)," ITU-T Recommendation X.1500, 2011.

25 The Internet Engineering Task Force, "An Incident Object Description Exchange Format (IODEF) Extension for Structured Cybersecurity Information," RFC 7203, April 2014.

26 International Telecommunications Union, "Discovery mechanisms in the exchange of cybersecurity information," ITU-T Recommendation X.1570, 2011.

Takeshi TAKAHASHI, Ph.D.

Senior Researcher, Cybersecurity Laboratory,
Cybersecurity Research Institute
Cybersecurity, Network Security