Advance Access publication on 8 October 2014
doi:10.1093/comjnl/bxu101

Reference Ontology for Cybersecurity Operational Information

Takeshi Takahashi[1,*] and Youki Kadobayashi[2]

[1]*Network Security Research Institute, National Institute of Information and Communications Technology, Tokyo 184-8795, Japan*
[2]*Graduate School of Information Science, Nara Institute of Science and Technology, Nara 8916-5, Japan*
[*]*Corresponding author: takeshi_takahashi@ieee.org*

As our cyber society develops and expands, the importance of cybersecurity operations is growing in response to cybersecurity threats coming from beyond national borders. Efficient cybersecurity operations require information exchanges that go beyond organizational borders. Various industry specifications defining information schemata for such exchanges are thus emerging. These specifications, however, define their own schemata since their objectives and the types of information they deal with differ, and desirable schemata differ depending on the purposes. They need to be organized and orchestrated so that individual organizations can fully exchange information and collaborate with one another. To establish the foundations of such orchestration and facilitate information exchanges, this paper proposes a *reference ontology* for cybersecurity operational information. The ontology structures cybersecurity information and orchestrates industry specifications. We built it from the standpoint of cybersecurity operations in close collaboration with cybersecurity organizations including security operation centers handling actual cybersecurity operations in the USA, Japan and South Korea. This paper demonstrates its usability by discussing the coverage of industry specifications. It then defines an extensible information structure that collaborates with such specifications by using the ontology and describes a prototype cybersecurity knowledge base we constructed that facilitates cybersecurity information exchanges among various parties. Finally, it discusses the usage scenarios of the ontology and knowledge base in cybersecurity operations. Through this work, we wish to contribute to the advancement of cybersecurity information exchanges.

Keywords: cybersecurity; ontology design; information structure; knowledge model; information schema; knowledge base; information exchange

Received 26 January 2014; revised 9 June 2014
Handling editor: David Rosado

1. INTRODUCTION

The widespread proliferation of the Internet is bolstering the development of a cyber society, in which diverse communications, including the sharing of private information and business transactions, are taking place. Nevertheless, it has also increased the number of cyber threats and diversified their targets and objectives [1, 2]. The targets range from individuals to private companies and even critical infrastructures such as nuclear power plants, whereas the objectives range from monetary benefit to political actions [3]. Accordingly, the need for cybersecurity operations is increasing in order to mitigate these threats.

In a cyber society, malware such as viruses may attack any computer beyond the borders of the country of its origin or target, and an attacker can attack computers all over the world by running other hackers' pre-packaged attack software. Sources of threats cross borders of countries and even continents, and an attacker can attack computers in country A by controlling computers in country B while physically residing in country C. Moreover, a system's vulnerability may be exposed to attackers across the globe. Countermeasures against these cybersecurity threats, however, are most frequently implemented by individual organizations in isolation. Consequently, an organization in one country may be attacked by malware whose

countermeasures are already known and implemented elsewhere. Such incidents occur because of a lack of information exchanges among organizations. Although some individual cybersecurity operators do exchange information locally, the primary methods are still e-mail, phone calls and even face-to-face meetings, which are not efficient.

To address this issue, various organizations have started to build information formats for sharing information beyond organization borders. To make these formats globally common, they are built in the forms of industry specifications or global standards. Such globally common formats provide two major advantages. First, they reduce the disparity of information availability on a global scale. From a technological standpoint, worldwide information sharing means that no country or organization is left behind in terms of information availability. Developing countries, which currently have fewer resources to put toward cybersecurity, can become equal partners with developed countries with appropriate investments; therefore, countermeasures can be implemented via global collaboration. Secondly, such formats streamline cybersecurity operations. Human operators often handle the operations manually since cybersecurity information in each organization is often not well structured. The formats structure information and make it machine-readable, thus facilitating information and knowledge management [4] and streamlining many of the operations.

1.1. Need for ontology building

Various industry specifications defining information schemata have already been built for sharing cybersecurity information among organizations [5, 6] (see Section 2.1 for details). They are useful for exchanging information for specific purposes, and parties can exchange information in a specific schema that they have agreed to use prior to the exchange. Nevertheless, it is difficult for them to exchange information in other schemata. Moreover, they may not find a suitable schema for exchanging information since existing specifications may not cover a sufficient range of information types and use cases. Thus, cybersecurity information exchanges among organizations and their automation still face difficulties in reality. Currently, there is no basis for determining their applicability, coverage and effectiveness. We need to take a holistic view of what types of information are needed and should be exchanged for maintaining cybersecurity.

To address this issue, we take an approach that considers who uses what types of information for what purposes and build an ontology of cybersecurity operational information. An ontology is an explicit specification of a conceptualization, which is an abstract and simplified view of the world that we wish to represent for particular purposes [7]. It structures information, serves as a basis of a knowledge architecture, and assists sharing and reutilization of knowledge [8]. An ontology of cybersecurity operational information can thus provide a framework for sharing and reusing such information and define the terminology. It can also orchestrate industry specifications for cybersecurity information schemata and facilitate discussion of their applicability, coverage and effectiveness. Several ontologies [9–14] (see Section 2.2 for details) have been developed for information security-related purposes. Although they were well formed and can be adapted over time to represent rapidly changing situations, they were built for different scopes and objectives; thus, they were not sufficient for our purposes.

1.2. Contribution

This paper proposes a *reference ontology* for cybersecurity operational information in order to build a basis for cybersecurity information exchange on a global scale. The ontology structures cybersecurity information, orchestrates and collaborates with industry specifications, and thus facilitates the exchange of an assortment of cybersecurity information in different schemata. This ontology, unlike others, has been developed in close collaboration with cybersecurity organizations, including security operation centers (SOCs) in the USA, Japan and South Korea. Although each cybersecurity organization runs slightly different operations, we succeeded in building a generalized ontology of cybersecurity operational information.

To demonstrate the ontology's usability, this paper reviews existing industry specifications of cybersecurity information schemata by mapping the specifications for each of the information types defined by the ontology. It also defines an extensible information structure that incorporates assorted cybersecurity information schemata by using the ontology; the structure becomes the basis for information sharing beyond organization borders. Building upon the information structure, this paper also introduces a cybersecurity knowledge base that organizes and accumulates cybersecurity information; its prototype can accumulate and retrieve assorted cybersecurity information in differing schemata. It finally discusses the usage scenarios of the ontology and knowledge base in cybersecurity operations. In doing this, we wish to contribute to the advancement of cybersecurity information exchanges.

1.3. Organization

The remainder of this paper is organized as follows. Section 2 introduces related work, Section 3 describes the methodology of building the proposed ontology and Section 4 introduces and elaborates on the ontology. Section 5 formalizes the ontology. To demonstrate the ontology's usability and applicability, Section 6 reviews existing industry specifications defining cybersecurity-related information schemata, defines an extensible information structure that collaborates with industry specifications, introduces a cybersecurity knowledge

TABLE 1. Industry specifications.

Specification name	Abbreviation	Organization	References
Asset Reporting Format	ARF	NIST	[15]
Common Attack Pattern Enumeration and Classification	CAPEC	ITU-T	[16]
Common Configuration Enumeration	CCE	NIST	[17]
Common Configuration Scoring System	CCSS	NIST	[18]
Common Event Expression	CEE	MITRE	[19]
Common Platform Enumeration	CPE	NIST	[20]
Common Result Format	CRF	MITRE	[21]
Common Vulnerabilities and Exposures	CVE	ITU-T	[22]
Common Vulnerability Reporting Framework	CVRF	ICASI	[23]
Common Vulnerability Scoring System	CVSS	ITU-T	[24]
Common Weakness Enumeration	CWE	ITU-T	[25]
Common Weakness Scoring System	CWSS	MITRE	[26]
Cyber Observable eXpression	CybOX	MITRE	[27]
Incident Object Description Exchange Format	IODEF	IETF	[28]
Malware Attribute Enumeration and Characterization	MAEC	ITU-T	[29]
Malware Metadata Exchange Format	MMDEF	IEEE	[30]
Open Checklist Interactive Language	OCIL	NIST	[31]
Open Vulnerability and Assessment Language	OVAL	ITU-T	[32]
Software Identification	SWID	ISO/IEC	[33]
Web Services Agreement Specification	WS-Agreement	Open Grid Forum	[34]
eXtensible Access Control Markup Language	XACML	OASIS	[35]
eXtensible Configuration Checklist Description Format	XCCDF	ISO/IEC	[36]

base that uses the information structure, and then discusses the usage scenarios of the ontology and knowledge base. Finally, Section 7 concludes the paper.

2. RELATED WORK

Various studies that aimed at exchanging cybersecurity-related information among parties have been reported. These include industry specifications defining information schemata and cybersecurity-related ontologies. There are also many guidelines that define common vocabulary and frameworks to be shared among parties. Cybersecurity information repositories are also available online, and they can share a variety of information with many parties. Although the proposed ontology is not particularly based on any of these works, they were useful in defining the ontology and provided a basis for its consideration.

2.1. Information schemata

Cybersecurity information needs to be machine-readable to enable efficient information exchange, retrieval and operation automation. Various industry specifications defining structures of cybersecurity information have been built to address this issue. Table 1 lists major ones including work-in-progress ones.[1] For instance, CVE defines the naming rule of identifiers to identify vulnerability information and an XML schema to describe vulnerability information,[2] IODEF defines an XML schema to describe incident information and XCCDF defines an XML schema to describe a checklist of security configurations. As a result, various cybersecurity information is expressed in differing schemata since the desirable schema for each kind of information differs depending on the information's usage purpose. To retrieve information in differing schemata, ordinary XML-based retrieval techniques [37–39] cannot be simply used. To achieve cybersecurity information retrieval, the gaps among these schemata must be considered.

One straightforward approach for this issue is to build a universal schema for all types of cybersecurity information. It is, however, very difficult to build such a schema since the desired schema depends on the information's usage. Having such a schema might even hinder the development

[1]The abbreviations of organizations in Table 1 are as follows. FIRST, the Forum of Incident Response and Security Teams; ICASI, the Industry Consortium for Advancement of Security on the Internet; IEEE, the Institute of Electrical and Electronics Engineers; IETF, the Internet Engineering Task Force; OASIS, the Organization for the Advancement of Structured Information Standards; NIST, the National Institute of Standards and Technology.

[2]CVE was originally designed to build a dictionary of vulnerabilities, and its XML schema is an extra feature added in response to users' requests.

of automation of cybersecurity operations. Moreover, even if such a schema were built, it might be subject to change since security techniques and operations will be further improved. Thus, such a schema might not be used in the future.

Instead of defining a universal schema, we need a design that can flexibly support and incorporate a variety of schemata in order to maintain the usability of cybersecurity information. Moreover, more industry specifications will emerge in the future, and they need to be identified and located. The number of structured cybersecurity information schemata is not that large at present, but it is expected to increase. An extensible information structure that can incorporate future industry specifications is thus needed. The ontology proposed in this paper can become the basis for such an information structure and can flexibly incorporate assorted schemata.

2.2. Ontologies

As discussed in Section 1.1, an ontology is an explicit specification of a conceptualization, which is an abstract, simplified view of the world that we wish to represent for particular purposes [7]. Ontologies are useful as means to support knowledge sharing and reutilization [8]. This reusability approach is based on the assumption that if a modeling scheme, i.e. an ontology, is explicitly specified and mutually agreed upon by the parties involved, then it is possible to share, reuse and extend knowledge.

Various work on ontologies in the area of cybersecurity has been reported. Fenz and Ekelhart [11] proposed a security ontology aimed at organizing knowledge of information security-related concepts with its focus on information security risk management based on their literature review. The ontology defines three sub-ontologies—security, enterprise and location—to organize concepts, and a description logic [40] is used to formalize the ontology. Wang *et al.* introduced a vulnerability ontology [13, 14]. It is designed for vulnerability analysis and management and captures the relationships among information technology (IT) products, vulnerabilities, attackers, security metrics, countermeasures and other relevant concepts. Tsoumas and Gritzalis [9] built an ontology of security management within an organization, with a focus on risk assessment. They built the ontology by extending the Distributed Management Task Force Common Information Model [41] with ontological semantics and converting it into Web Ontology Language (OWL) [42]. They also provided a framework that uses the ontology for policy-based risk assessment at the concept level. Parkin *et al.* [12] proposed an information security ontology incorporating human-behavioral implications. This ontology provides a framework for investigating casual relationships of human-behavioral implications resulting from information security management decisions before security controls are deployed. Denker *et al.* [10] proposed several ontologies for security annotations of agents and web services using OWL. They mainly addressed knowledge representation and some of the reasoning issues for trust and security in the Semantic Web. Masoumzadeh and Joshi [43] introduced an ontology for Social Networking Systems (SNSs), which captures privacy-sensitive information in SNSs. They used the ontology to discover missing privacy protection policies in SNSs. Although there have been various other ontology studies [44], the reusability of those ontologies for our purpose is rather limited, or they are still at early stages of development.

Unlike the aforementioned work, our ontology is designed for actual cybersecurity operations and focuses on cybersecurity operational information. For practicality and reusability, we build it on the basis of intensive discussions with cybersecurity operators. This ontology can provide a framework for sharing and reusing cybersecurity operational information and can define the terminology.

2.3. Cybersecurity guidelines

When building an ontology, cybersecurity guidelines are useful for understanding various aspects of cybersecurity operations and finding terminologies.

Assorted international standards bodies have produced such guidelines. ISO/IEC 27032 [45] provides guidelines for cybersecurity; it outlines assorted cybersecurity concepts and technical controls and provides guidelines for information sharing and coordination. ITU-T Recommendation E.409 [46] describes incident handling operations, while ITU-T Recommendation X.1500 [47, 48] provides an overview of cybersecurity information exchange. IETF Request for Comments 2350 [49] describes the general expectations of Computer Security Internet Response Teams that run incident response operations.

In addition to these international standards bodies, various other organizations have produced cybersecurity guidelines. NIST's Special Publications in the 800 series [50] present wide-ranging cybersecurity guidelines, e.g. on basic ideas for computer security [51], security services [52], incident handling [53], forensics [54], testing [55] and measurement [56]. Control Objectives for Information and Related Technology [57] is a set of best practices that provides a framework for the governance and management of enterprise IT; it indicates the operations that are necessary for organizations to maintain cybersecurity. The Framework for Improving Critical Infrastructure Cybersecurity [58], which was built in response to Executive Order 13636 [59] issued by President Obama, provides a framework for critical infrastructure in order to maintain its cybersecurity; it can be used as a key part of an organization's systematic process for identifying, assessing and managing cybersecurity risk. It can be used not only by critical infrastructure organizations but also other organizations to maintain their cybersecurity.

Our ontology was built based on intensive discussions we held with cybersecurity operators, and these guidelines helped

us to carry out deeper discussions and a more thorough analysis of actual cybersecurity operations. Indeed, these guidelines were also built based on the knowledge of industry experts including cybersecurity operators. Thus, studying these guidelines was an indispensable step in creating an ontology that was in alignment with actual cybersecurity operations. Moreover, the guidelines facilitated our work by providing a common vocabulary. The existing vocabularies were used differently by each organization, so the guidelines worked as a common ground for discussion.

Along with the work on cybersecurity guidelines, there are also studies in the area of policies and law. An OECD report [3] summarizes the recent trend of cybersecurity policy making of governments. Likewise, the Tallinn Manual on the International Law Applicable to Cyber Warfare [60] examines laws applicable to cyber war. Although national policies and laws themselves are outside the scope of this paper, they are closely related to cybersecurity and provided useful information in relation to cybersecurity operations.

2.4. Public repositories

As attempts to share and circulate security-related information, there are several online repositories that are publicly available. The National Vulnerability Database (NVD) [61] is a repository of vulnerability information. Each piece of vulnerability information has a CVE identifier (CVE ID), and its data structure conforms to CVE with minor extensions. Open Sourced Vulnerability Database (OSVDB) [62] is another vulnerability database that is independent, open-source and web-based. It provides vulnerability information on the web and assigns its own identifiers to each item of information with a note of relevant CVE IDs. Red Hat provides repositories for CVRF-based vulnerability information and OVAL-based security check information [63]. Japan Vulnerability Notes (JVN) [64] provides vulnerability information in Japanese and is described in accordance with its own schema in Resource Description Framework (RDF) [65]. MITRE also provides several repositories including a CVE-based one and an OVAL-based one [6], Many more such repositories will be provided by organizations around the world in the future.

Although assorted information will be available online, these repositories may accumulate information using differing schemata. The schema gap needs to be addressed to accumulate comprehensive information. The cybersecurity knowledge base built on the proposed ontology, introduced in this paper, copes with the schema gap and accumulates cybersecurity information in differing schemata.

3. METHODOLOGY

In order to build a practical ontology, we collaborated with nine major cybersecurity organizations including three SOCs that handle actual cybersecurity operations in the USA, Japan and South Korea, and we designed the ontology for collaboration with industry specifications.

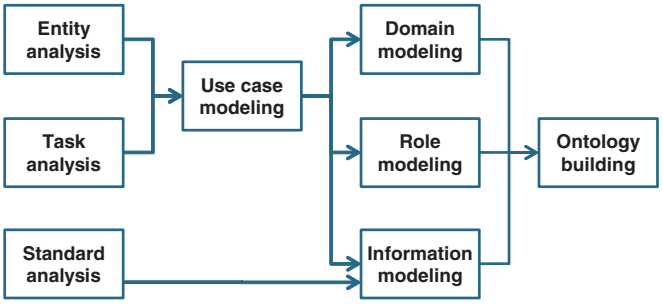

FIGURE 1. Methodology.

The methodology we used in building the ontology is shown in Fig. 1. The solid arrows show the order of the work flow. We used a software engineering and business process modeling approach and built several models in Unified Modeling Language and then distilled them into the proposed ontology. First, we conducted an entity analysis and task analysis based on intensive discussions with the SOCs. During this process, we dealt with confidential information and other sensitive details of those centers by working individually with them in separate sessions. We used the analyses to conduct use case modeling, where we generalized the analysis results in order to disregard differences in individual organizations' operations and ensure that sensitive information would be hidden. In parallel, we conducted an analysis of existing standards. Based on the use case modeling and standard analysis, we conducted domain, role and information modeling, with which we built the ontology. While building the models and ontology, we iterated the process of discussion and review with the cybersecurity organizations, and ultimately succeeded in building a generalized ontology of cybersecurity operational information.

Note that we have conducted an extensive literature review that is not depicted in Fig. 1, but it became the basis for considering each stage of this work.

4. PROPOSED ONTOLOGY

Following the above methodology, this section proposes a *reference ontology* for cybersecurity operational information, which is depicted in Fig. 2. It consists of operation domains, roles required to run operations in the domains and cybersecurity information associated with the roles. It thus depicts *who* (*role*) *uses what type of information* (*cybersecurity information*) *for what purpose* (*operation domain*). This section elaborates on the operation domains, roles and cybersecurity information.

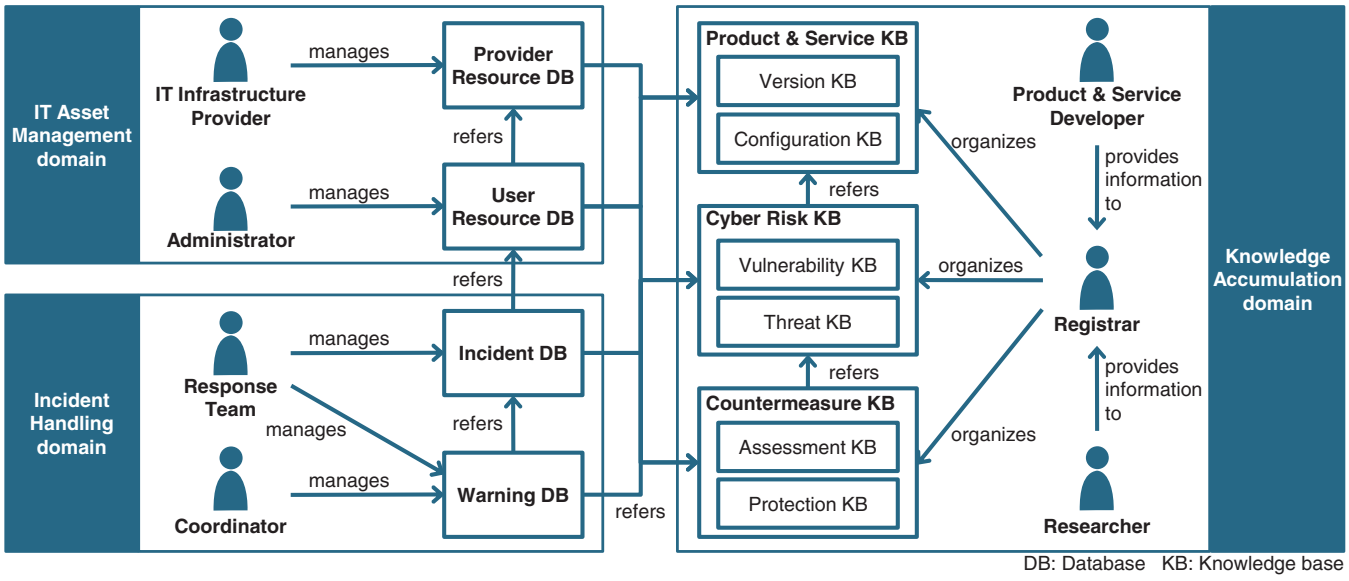

FIGURE 2. Ontology of cybersecurity operational information.

4.1. Operation domains

The term 'cybersecurity operation' covers a range of security operations in cyber society, but this paper focuses on cybersecurity operations that preserve information security in cyber societies. Information security is the preservation of information confidentiality, integrity and availability [66]. It sometimes also encompasses non-repudiation, accountability, authenticity and reliability of information [67]. To represent the domains of such operations, the proposed ontology defines three operation domains: IT Asset Management, Incident Handling and Knowledge Accumulation.

IT asset management runs cybersecurity operations inside user organizations such as installing, configuring and managing IT assets, and it covers both incident prevention and damage control operations. IT assets include not only a user's own IT assets but also network connectivity, cloud services and identity services provided by external entities for the user.

Incident handling detects and responds to incidents occurring in cyber societies by monitoring computer events, incidents comprising multiple computer events and attack behaviors that caused the incidents. More specifically, it monitors computer events, and when an anomaly is detected, it produces an incident report. Based on the report, it investigates the incident in detail so that it can clarify the attack pattern and its countermeasures. Based on the incident analysis, it may provide alerts and advisories, e.g. early warnings against potential threats, to user organizations.

Knowledge accumulation collects and generates cybersecurity information and extracts reusable knowledge for other organizations. To facilitate the reusability, it provides common naming and taxonomy, with which it organizes and accumulates the knowledge. This domain serves as the basis of global collaboration beyond organization borders.

4.2. Role

Based on the operation domains defined in Section 4.1, this section identifies roles necessary for running cybersecurity operations in each domain. These roles are listed in Table 2. The IT Asset Management domain has Administrator and IT Infrastructure Provider, the Incident Handling domain has Response Team and Coordinator and the Knowledge Accumulation domain has Researcher, Product & Service Developer and Registrar. Note that the roles are defined from the viewpoint of functions; therefore, one entity may perform several roles depending on the context.[3]

Administrator administers its organization's system and maintains its functionality. For this purpose, this role monitors system usage, diagnoses the system by running integrity checks, scanning for vulnerabilities and running penetration tests, and then assesses the system's security level. A system administrator inside each organization is a typical instance. A Managed Security Service Provider (MSSP) also serves as an Administrator if an organization outsources some of the above operations.

[3] A critical infrastructure organization may serve as an Administrator or IT Infrastructure Provider, depending on the context. If the context is the protection of the organization, the organization (or its administrators) serves as Administrator and runs cybersecurity operations to protect itself, as with any other organization. If the context is the protection of other organizations, critical infrastructure organizations serve as the IT Infrastructure Provider; we can see that individual organizations use services provided by critical infrastructure organizations.

TABLE 2. Operation domains and roles.

Operation domains	Roles
IT Asset Management	Administrator
	IT Infrastructure Provider
Incident Handling	Coordinator
	Response Team
Knowledge Accumulation	Product & Service Developer
	Registrar
	Researcher

IT infrastructure provider provides the appropriate IT infrastructure, including resources and services, to an organization. The infrastructure includes the network connectivity and cloud services such as software as a service (SaaS), platform as a service and infrastructure as a service. This role maintains the quality and security of the infrastructure so that user organizations can enjoy the best of it. For instance, it implements access control, monitors access logs and controls traffic flow on the network. An Internet service provider, application service provider and cloud service provider are typical instances.

Response team monitors and analyzes events in an organization. It detects incidents, e.g. unauthorized access, distributed denial of service (DDoS) attacks and phishing, and accumulates incident information. It also runs triage (or sometimes remediation) on the incident by collaborating with Administrator or IT Infrastructure Provider. For instance, it may ask an administrator of a user organization to unplug its computers from networks, or it may ask a network provider to register phishing site addresses on its blacklists or block malicious traffic. The incident response team inside an MSSP is a typical instance. In many user organizations, system administrators often work as not only Administrator but also Response Team.

Coordinator coordinates with the other roles and addresses potential threats based on known incidents and crime information. It provides warnings to other organizations and sometimes leads the collaborative mitigation to handle devastating and large-scale attacks such as DDoS attacks. The collaboration between Response Team, Administrator and IT Infrastructure Provider often requires coordination provided by the Coordinator, if these roles belong to different organizations. The CERT Coordination Center (CERT/CC), be it either commercial or non-commercial, is a typical instance.

Researcher researches cybersecurity issues including vulnerabilities and attacks, extracts knowledge from the research and accumulates the knowledge. It publishes a lot of reusable information through Registrar so that individual organizations may implement needed countermeasures. X-force within International Business Machines Corp. (IBM), the Risk Research Institute of Cyber Space at the Little eArth Corporation Co., Ltd. (LAC) and McAfee Lab within McAfee, Inc. are typical instances.

Product & Service Developer develops products and services and accumulates information about them, such as their versions, configurations, vulnerabilities and patches. It publishes a lot of reusable information through Registrar so that, as with Researcher, individual organizations may implement needed countermeasures. A software vendor and individual private software programmer are typical instances.

Registrar classifies, organizes and accumulates cybersecurity knowledge provided by the Researcher and Product & Service Developer so that the knowledge can be reused by other organizations. NIST and the IT Promotion Agency, Japan are typical instances. In some cases, an entity serving as Researcher or Product & Service Developer may also serve as Registrar and publish information.

4.3. Cybersecurity information

Based on the operation domains and roles, this section identifies cybersecurity information needed for operations. Considering the information with which each of the roles is involved, we define four databases—User Resource, Provider Resource, Incident and Warning—and three knowledge bases: Product & Service, Cyber Risk and Countermeasure. Note that both a database and knowledge base accumulate information, but we use these terms differently; most of the information in a database is not refined enough to be shared with and reused by other organizations, whereas most of the information in a knowledge base is sufficiently refined to be shared with and reused by other organizations.

4.3.1. User Resource Database

This database accumulates information on assets inside an organization. The information it contains typically consists of lists of software/hardware, their configurations, resource usage status, security level assessment results, Intranet topology, data provenance [68], information security policies including access control policies and standards and guidelines that the organization follows. It also contains external resource information that the user organization uses such as lists of subscribed online services (e.g. data centers and SaaS) and their usage records. Administrator manages such information. ARF can be used for describing the IT assets within an organization, XACML can be used for describing access control policies, while CVSS/CWSS can be used for scoring the security level of the IT asset.[4] The scores are useful for Administrators in prioritizing the urgency of security operations on IT assets.

[4]Note that specification conformance is not strictly required for streamlining cybersecurity operation—organizations may employ different standards for asset enumeration, for instance. The ontology gives us essential structure for linking cybersecurity information, as well as choice of standards that organizations may adopt in a long term.

4.3.2. Provider Resource Database

This database accumulates information that is necessary for user organizations to run cybersecurity operations and that belongs to and is managed by the IT Infrastructure Provider. The database mainly contains information on networks, server assets and policies. Network information concerns networks with which each user organization is connected, such as topology, routing information, access control policies, traffic status and packet logs. Server asset information includes the access logs, service usage records, anomaly detection reports and workload information. Policy information includes terms and conditions, service specifications, service level agreements, the information security policy and standards and guidelines that the IT Infrastructure Provider follows. WS-Agreement can be used for describing the service agreement. Note that user organization specific information such as the local configuration of each cloud service is stored in the User Resource Database. In order to run effective and efficient cybersecurity operations, the database needs to be linked to a User Resource Database. The border between internal and external IT assets thus becomes increasingly unclear, especially in cloud computing.

4.3.3. Incident Database

This database contains information on incidents, which is generated mainly from an analysis of information in the User Resource Database.[5] The Response Team manages the information. This database includes three records: Event Record, Incident Record and Attack Record.

Event record contains information on computer events including that on packets, files and their transactions. Usually computers automatically provide most of the records as computer logs, such as for log-in time and date as well as terminal information provided when root users log in to a system. The logs are instances of this record. CEE and CybOX can be used to describe the record.

Incident record contains information on security incidents and provides information such as the current state of user systems and further risks. It is derived from analyses of several Event Records and their conjectures, which are created automatically or manually. For instance, when excessive access to one computer is detected, the state of the computer (excessive access to one computer) and its expected consequence (denial of service) should be recorded in the Incident Record. The extent of the harm caused by the incident as well as the need for countermeasures can be judged from this record in accordance with the information security policy, standards and guidelines. Note that an Incident Record may record false incidents; i.e. incident candidates judged as non-incidents after an investigation. IODEF can be used to describe the record.

Attack record contains information on attacks derived from analyses of Incident Records. It describes the attack sequence; such as how the attack was initiated, which IT assets were targeted, and how the attack's damage propagated. Note that this record needs to be linked to the Incident Record.

4.3.4. Warning Database

This database contains information on cybersecurity warnings. The information is designed for either the general public or specific organizations. The information for the general public usually consists of statistical information and alerts, while the information for specific organizations consists of the security policy and guidelines as well as security advice customized for the organization. The information is generated mainly from information in the Incident Database and Cyber Risk Knowledge Base. The Coordinator and Response Team manage such information. Based on the warnings, user organizations may implement countermeasures against warned cybersecurity risks.

4.3.5. Product & Service Knowledge Base

This knowledge base accumulates information on products and services. It is provided by the Researcher and Product & Service Developer, and is then organized and classified by the Registrar. It includes the Version Knowledge Base and Configuration Knowledge Base.

Version knowledge base accumulates version information on products and services, which includes naming and enumeration of their versions. Security patches of software products are also included here. CPE identifiers and SWID tags can be used to enumerate software assets and platforms.

Configuration knowledge base accumulates configuration information about products and services. It includes naming, taxonomy and enumeration of known configurations of products and services. Regarding service configuration, it also contains guidelines for service usages. CCE can be used to enumerate common configurations of products.

4.3.6. Cyber Risk Knowledge Base

This knowledge base accumulates cybersecurity risk information. It is provided by the Researcher and Product & Service Developer, and is then organized and classified by the Registrar, as with the other knowledge bases. It includes the Vulnerability Knowledge Base and Threat Knowledge Base.

Vulnerability knowledge base accumulates known vulnerability information, which includes naming, taxonomy and enumeration of known software and system vulnerability. The vulnerability information covers the vulnerabilities caused by both programming and configuration. It also includes information on human vulnerabilities, which are vulnerabilities to

[5] An incident response operation sometimes requires information from the IT Infrastructure Provider, but the information does not come to the response team directly but comes through the user organization. To reflect this, we modeled information needed for the incident response being extracted from the Provider Resource Database and entered into the User Resource Database and then extracted from the User Resource Database and entered into the Incident Database.

which human IT users are exposed. NVD and OSVDB are practical instances of this database. CVE and CWE can be used to describe the contents of the knowledge base.

Threat knowledge base accumulates known cybersecurity threat information. It has knowledge of attacks and misuses. Attack knowledge includes attack patterns, attack tools (e.g. malware), and their trends (e.g. statistical information on attacks in terms of geography, target organization types and exploited vulnerabilities). CAPEC and MAEC can be used to describe the knowledge. Misuse knowledge includes information on misuses attributed to users' inappropriate usages, whether benign or malicious. Benign usages include mistyping, misrecognition caused by inattentional blindness [69], misunderstanding and being caught in phishing traps, whereas malicious usages include compliance violations such as unauthorized service usage and access to inappropriate materials.

4.3.7. *Countermeasure Knowledge Base*

This knowledge base accumulates information on countermeasures to cybersecurity risks. It is provided by the Researcher and Product & Service Developer, and is then organized and classified by the Registrar, as with the other knowledge bases. It has the Assessment Knowledge Base and Protection Knowledge Base.

Assessment knowledge base accumulates known rules and criteria for assessing the security level of IT assets, checklists of configurations and heuristics including best practices. CCSS, CVSS and CWSS provide formulas for assessing security levels, and the assessment results that might be reusable by other organizations (e.g. vulnerability severity scores) are accumulated in this knowledge base. XCCDF and OVAL can be used to describe rules and provide checklists, and their scripts are also accumulated in this knowledge base.

Protection knowledge base accumulates known information on detecting and preventing security threats. It includes blacklist URLs and the list of open resolvers and email servers allowing third-party email relay. It also includes signatures of intrusion detection systems and intrusion prevention systems and detection/protection rules that follow the signatures. It also accumulates heuristics including best practices.

5. FORMALIZATION

This section formalizes the proposed ontology to reduce ambiguity. Although there are ontology editors and tools available for this purpose, this section uses description logic [40]. Table 3 shows the list of concepts defined in Section 4 and their abbreviations, which will be used in this section to describe the relations between concepts with description logic.

The ontology defines three operation domains $\mathrm{OD} = \{\mathrm{OD_{Asset}}, \mathrm{OD_{Incdt}}, \mathrm{OD_{Knwl}}\}$ and seven roles $\mathrm{RL} = \{\mathrm{RL_{Adm}}, \mathrm{RL_{Prvdr}}, \mathrm{RL_{Coord}}, \mathrm{RL_{Resp}}, \mathrm{RL_{Rgstr}}, \mathrm{RL_{Dev}}, \mathrm{RL_{Rsr}}\}$. The roles are necessary to run operations. The relations between them are formalized as below.

$$\{\mathrm{RL_{Adm}}, \mathrm{RL_{Prvdr}}\} \sqsubseteq \exists \mathrm{handlesOperations}.\mathrm{OD_{Asset}},$$
$$\{\mathrm{RL_{Coord}}, \mathrm{RL_{Resp}}\} \sqsubseteq \exists \mathrm{handlesOperations}.\mathrm{OD_{Incdt}},$$
$$\{\mathrm{RL_{Rgstr}}, \mathrm{RL_{Dev}}, \mathrm{RL_{Rsr}}\} \sqsubseteq \exists \mathrm{handlesOperations}.\mathrm{OD_{Knwl}}.$$

The roles use $\mathrm{CI} = \{\mathrm{KB_{CM}}, \mathrm{KB_{Risk}}, \mathrm{KB_{Incdt}}, \mathrm{KB_{P\&S}}, \mathrm{DB_{Prvdr}}, \mathrm{DB_{User}}, \mathrm{DB_{Warn}}\}$ to run their operations. The relations between the roles and cybersecurity information are formalized as below.

$$\begin{aligned}
\mathrm{RL_{Prvdr}} &\sqsubseteq \exists \mathrm{managesDB}.\mathrm{DB_{Prvdr}},\\
\mathrm{RL_{Adm}} &\sqsubseteq \exists \mathrm{managesDB}.\mathrm{DB_{User}},\\
\mathrm{RL_{Resp}} &\sqsubseteq \exists \mathrm{managesDB}.\mathrm{DB_{Incdt}}\\
&\quad \sqcap \exists \mathrm{managesDB}.\mathrm{DB_{Warn}},\\
\mathrm{RL_{Coord}} &\sqsubseteq \exists \mathrm{managesDB}.\mathrm{DB_{Warn}},\\
\mathrm{RL_{Rgstr}} &\sqsubseteq \exists \mathrm{managesKB}.\mathrm{KB_{P\&S}}\\
&\quad \sqcap \exists \mathrm{managesKB}.\mathrm{KB_{CM}}\\
&\quad \sqcap \exists \mathrm{managesKB}.\mathrm{KB_{Risk}}.
\end{aligned}$$

TABLE 3. List of concepts and abbreviations.

Concept	Abbreviation
Operation Domains	OD
IT Asset Management	$\mathrm{OD_{Asset}}$
Incident Handling	$\mathrm{OD_{Incdt}}$
Knowledge Accumulation	$\mathrm{OD_{Knwl}}$
Roles	RL
Administrator	$\mathrm{RL_{Adm}}$
Coordinator	$\mathrm{RL_{Coord}}$
IT Infrastructure Provider	$\mathrm{RL_{Prvdr}}$
Product & Service Developer	$\mathrm{RL_{Dev}}$
Registrar	$\mathrm{RL_{Rgstr}}$
Researcher	$\mathrm{RL_{Rsr}}$
Response Team	$\mathrm{RL_{Resp}}$
Cybersecurity Information	CI
Provider Resource DB	$\mathrm{DB_{Prvdr}}$
User Resource DB	$\mathrm{DB_{User}}$
Incident DB	$\mathrm{DB_{Incdt}}$
Warning DB	$\mathrm{DB_{Warn}}$
Countermeasure KB	$\mathrm{KB_{CM}}$
Assessment KB	$\mathrm{KB_{Assmt}}$
Protection KB	$\mathrm{KB_{Protn}}$
Cyber Risk KB	$\mathrm{KB_{Risk}}$
Vulnerability KB	$\mathrm{KB_{Vuln}}$
Threat KB	$\mathrm{KB_{Thrt}}$
Product & Service KB	$\mathrm{KB_{P\&S}}$
Version KB	$\mathrm{KB_{Ver}}$
Configuration KB	$\mathrm{KB_{Cfg}}$

DB, Database; KB, Knowledge Base.

The types of cybersecurity information CI are classified into databases DB and knowledge bases KB, where $\mathrm{DB} = \{\mathrm{DB_{User}}, \mathrm{DB_{Prvdr}}, \mathrm{DB_{incdt}}, \mathrm{DB_{Warn}}\}$ and $\mathrm{KB} = \{\mathrm{KB_{CM}}, \mathrm{KB_{Risk}}, \mathrm{KB_{P\&S}}\}$, and their relations are formalized as below.

$$\begin{aligned}
\mathrm{DB} &\sqsubseteq \exists \mathrm{refersKB.KB_{P\&S}} \sqcap \exists \mathrm{refersKB.KB_{Risk}} \\
&\quad \sqcap \exists \mathrm{refersKB.KB_{CM}}, \\
\mathrm{DB_{User}} &\sqsubseteq \exists \mathrm{refersDB.DB_{Prvdr}}, \\
\mathrm{DB_{Incdt}} &\sqsubseteq \exists \mathrm{refersDB.DB_{User}}, \\
\mathrm{DB_{Warn}} &\sqsubseteq \exists \mathrm{refersDB.DB_{Incdt}}, \\
\mathrm{KB_{Risk}} &\sqsubseteq \exists \mathrm{refersKB.KB_{P\&S}} \sqcap \exists \mathrm{hasKB.KB_{Vuln}} \\
&\quad \sqcap \exists \mathrm{hasKB.KB_{Thrt}}, \\
\mathrm{KB_{CM}} &\sqsubseteq \exists \mathrm{refersKB.KB_{Risk}} \sqcap \exists \mathrm{hasKB.KB_{Assmt}} \\
&\quad \sqcap \exists \mathrm{hasKB.KB_{Protn}}, \\
\mathrm{KB_{P\&S}} &\sqsubseteq \exists \mathrm{hasKB.KB_{Ver}} \sqcap \exists \mathrm{hasKB.KB_{Cfg}}.
\end{aligned}$$

There are also relations among roles. The ontology defines the minimum number such relations, which are formalized as below.

$$\{\mathrm{RL_{Dev}}, \mathrm{RL_{Rsr}}\} \sqsubseteq \exists \mathrm{providesInformationTo.RL_{Rgstr}}.$$

One could argue that more relations exist among the defined concepts; for instance, Response Team collaborates with Administrator with the assistance of Coordinator. Nevertheless, an ontology could be defined differently depending on the focus and purpose of its modeling, and the focus of this ontology modeling is cybersecurity information. Thus, the relations among roles and operation domains are kept to the minimum in this ontology.

Note that, for consistency checking, we converted the logics in this section into OWL Description Logic, used the inconsistency check function of Protégé [70], and confirmed the absence of errors.

6. USABILITY AND APPLICABILITY

This section demonstrates the ontology's usability and applicability. We review industry specifications based on the ontology, define an extensible information structure that incorporates assorted industry specifications, introduce a cybersecurity knowledge base that follows the information structure, and finally, discuss streamlining of cybersecurity operations.

6.1. Reviewing industry specifications

There are various specifications defining schemata of cybersecurity-related information, but their coverage needs to be reviewed. This section uses the ontology to review the coverage.

TABLE 4. Incorporating structured information.

Categories	Formats
User Resource DB	ARF, XACML
Provider Resource DB	WS-Agreement
Incident DB	CEE, CybOX
Warning DB	IODEF
Cyber Risk KB	
Vulnerability KB	CVE, CVRF, CWE
Threat KB	CAPEC, MAEC, MMDEF
Countermeasure KB	
Assessment KB	CCSS, CVSS, CWSS
Protection KB	OCIL, OVAL, XCCDF
Product & Service KB	
Version KB	CPE, SWID
Configuration KB	CCE

DB, Database; KB, Knowledge Base.

The specifications for the information types defined by the ontology are listed in Table 4. Note that a specification could be used by more than one information type, but for simplicity this table assigns each specification to one information type: the most suitable one for it. That suits the best for the specification for simplicity. For instance, IODEF could be used in the Incident Database and in the Warning Database, but we put it only in the Warning Database, since it is mainly used for information exchange among organizations and raw information on incidents inside an organization is not usually shared with any other organizations.

The table shows that there are more specifications for knowledge bases than for databases. This is natural since knowledge bases have information prepared for the purpose of sharing information beyond organization borders whereas databases have information used mainly inside an organization. Although the information inside a database needs to be shared to advance security operations, the industry is still at the stage of building specifications for knowledge bases.

The table also shows that each of the categories has at least one industry specification, but we do not believe we have sufficient specifications. By investigating each of the information types, we may find information sub-categories whose information needs to be structured and shared by parties. For instance, the Threat Knowledge Base contains attack and misuse knowledge, but no major industry specifications were found for describing misuse information. Likewise, the Version KB has CPE to define the naming of products but does not cover that of online services yet: this would, for instance, facilitate the listing of subscribing online services in the User Resource Database. Although further discussion may continue on this specification coverage issue, that is, outside the scope of this paper.

New industry specifications may emerge in the future, but their visibility and usability will be limited if they are scattered. The proposed ontology can clarify the position of each industry specification. It can also organize and orchestrate such specifications. As an example, Section 6.2 discusses one type of such orchestration. Note that ITU-T Recommendation X.1500, which describes the overview of cybersecurity information exchange, uses the proposed ontology in its appendix to organize and orchestrate industry specifications. Its appendix introduces various industry specifications. Since new industry specifications are continually emerging, the appendix is reviewed and revised periodically. The ontology is in another appendix and classifies and organizes the industry specifications so that it can provide the big picture of industry specifications and facilitate readers' understanding of these specifications. This also demonstrates the ontology's usability.

Along with the coverage of the schemata, their applicability and effectiveness should be reviewed. The ontology depicts who (role) uses what type of information (cybersecurity information) for what purpose (operation domain), as mentioned in Section 4. Thus, schemata can be reviewed to determine whether they are useful for the related roles and operation domains. For instance, we could review whether the IODEF schema is applicable and effective for the entities serving the roles of Coordinator and Response Team and running the Incident Handling operations. By reviewing the schema, we can see that it can convey assorted information, such as the incident ID, information on the systems involved in an event and contact information, that is, applicable in order to communicate the current security situation beyond organizational borders and that is effective for the current operations of several such entities. We can also see that the schema could be advanced further to accommodate future operations of some additional entities, especially the entities serving in the Response Team role. Indeed, the community is currently reviewing and revising the IODEF so that it can accommodate the most recent as well as future operations of the Response Team and Coordinator [71]. As with this example, the ontology can also be used to analyze the applicability and effectiveness of the schemata by considering who uses the schema for what purposes and operations.

6.2. Extensible information structure

An extensible information structure that can serve as a platform for incorporating cybersecurity information schemata is needed. It can become the basis for sharing assorted types of cybersecurity information among parties. This section uses the ontology to define such an information structure.

The proposed ontology serves as such an information structure basis since it defines a high-level taxonomy of cybersecurity information. Although a detailed taxonomy may help to define a specific format, it may produce an inflexible and unusable format since the desirable format depends on usage and may change in the future. Instead of defining a detailed taxonomy, we define an information structure that links the ontology and industry specifications defining information formats.

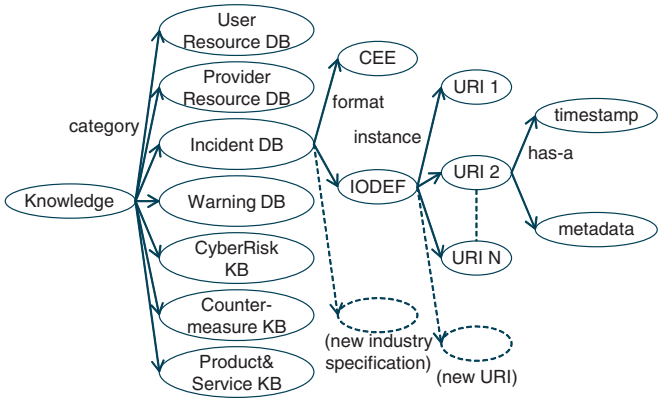

FIGURE 3. Information structure.

The information structure separates categories and formats; for these, it uses the cybersecurity information types defined by the ontology and industry specifications, respectively. Since the ontology can organize and orchestrate industry specifications, as we have seen in Table 4 in Section 6.1, the information structure links the types with the specifications according to the table. In this way, the specifications supplement the ontology by defining the detailed format of each information type defined by the ontology. Several industry specifications exist for differing purposes, and the information structure orchestrates them by using the ontology so that it can flexibly describe various cybersecurity information.

An overview of the information structure is shown in Fig. 3, where the information categories and formats are linked in accordance with the above discussion. Information entries expressed in a schema are registered under their corresponding industry specifications. The index of each information entry is the information's URI, and its metadata and timestamp follow the URI. The URI points to the information's location (e.g. file pointer, URL), the metadata outlines the information's content and is necessary to retrieve the information, and the timestamp records the last time the registry checked the information's existence. Note that, an industry specification may be used by multiple information types, as discussed in Section 6.1. Therefore, for instance, we may link IODEF with both the Incident Database and the Warning Database, as needed. (Indeed, Fig. 3 also links IODEF under the Incident DB.)

This information structure is extensible. First, the information format is extensible. We do not have sufficient specifications, as discussed in Section 6.1, and some information may not have proper specifications for describing its content. In this case, we simply need to build a new specification or extend

an existing one,[6] and the ontology simply uses such specifications as a means of describing the details of information types that it has defined. If the information structures defined by existing specifications become obsolete, we simply need to build a new specification and associate it with one of the categories specified by the ontology. Secondly, the categories are also extensible. New categories can be added without modifying any other part of the information structure, and arbitrary industry specifications, either new or existing, can be linked to them, if necessary. Nevertheless, we do not see the need for that at this stage since the underlying ontology was designed on the basis of the current operations of multiple major international SOCs through a year-long discussion and analysis (see Section 2.2), and the categories are abstract enough to absorb minor changes. In this way, the categories defined by the ontology are semi-fixed while the formats provided by various specifications are flexible, and the information structure is designed so that any changes in them could be easily made and propagated.

This information structure can be implemented in different ways, including an RDF-based implementation. RDF is a syntactic and semantic language for representing information describing available resources. It achieves the structure described in Fig. 3 by listing triples. RDF is designed to be extensible, so the extensibility discussed above is easily achieved; we can add an information entry, format, or even category by adding several lines at the end of the repository without needing to change existing entries inside the repository. The metadata of each information entry can be generated by running a predefined XSLT script on the information expressed in XML.

Note that the information structure uses industry specifications that define XML schemata of cybersecurity information; thus, we do not define any new formats but use industry specifications that define information formats with the assistance of the ontology. Application to the internal repository of Registries is described in the next section.

6.3. Cybersecurity knowledge base

Cybersecurity information needs to be shared among various parties to minimize security incidents. Building an online knowledge base is an efficient way to do this. This section introduces a prototype implementation of a cybersecurity knowledge base that uses the ontology-based information structure defined in Section 6.2.

The prototype organizes and accumulates XML-based cybersecurity information in accordance with the information structure. It accumulates links (URIs) to the locations of cybersecurity information in arbitrary schemata, metadata and timestamps. The metadata is generated by converting all the tags of the XML-based information into RDF through the use of XSLT. Although a meticulous metadata extraction mechanism could be implemented, the prototype was given this simple conversion for simplicity. It also lets users retrieve information accumulated within it through its web interface.

The prototype is implemented in Java on Linux CentOS. It uses Jena SDB [72], an implementation of SPARQL [73] engine, for its information retrieval functionality. For the purpose of demonstration, we prepared test data by copying entries of NVD and JVN as well as the CVE and OVAL repositories of MITRE and Red Hat (see Section 2.4 for details) and creating manually made test entries.

The web interface of the knowledge base is shown in Fig. 4. It provides four types of search interfaces, i.e. keyword search, tag-based search, category search and security information update. The keyword search is at the top of the figure. Users can enter an arbitrary keyword in the text box and run a search by clicking on the ‘Send Query’ button. They can perform more sophisticated searches by using a detailed search, which is found in the middle of the figure. They may specify the target tags of the retrieval. Note that they may lookup the available tags by clicking on the button located next to the text box. They can thus simply select a tag and then identify the keyword in the detailed search. The category-based search is in the lower left part of the figure. It provides a list of information categories, and users can select one of them to see the list of information entries in the entry. Users can see an entry’s content on the browser by clicking on one of the entries. The security information update is in the lower right part of the figure. It provides the latest cybersecurity information from information sources. Users can specify which information they are interested in and can filter the list of information displayed here, if necessary. Note that, the prototype runs tag-based searches internally by running the SPARQL engine regardless of which of the above interfaces is used.

The cybersecurity knowledge base is extensible since it uses the information structure defined in Section 6.2. It can thus support and incorporate XML-based information in a new schema just by correlating the schema with one of the categories and preparing appropriate XSLT scripts to extract metadata from the information. Therefore, the mechanism can instantly cope with new industry specifications defining new schemata.

6.4. Streamlining cybersecurity operations

The ontology facilitates structuring of cybersecurity information, thus streamlining the information management operations inside organizations. Information management is the basis of cybersecurity and cyberdefense. This section illustrates that by presenting a discussion on several usage scenarios we wish to realize with the ontology.

Administrators inside organizations can unify the management of various kinds of security information by using the

[6] These activities need to be initiated by users such as industry players or standards bodies.

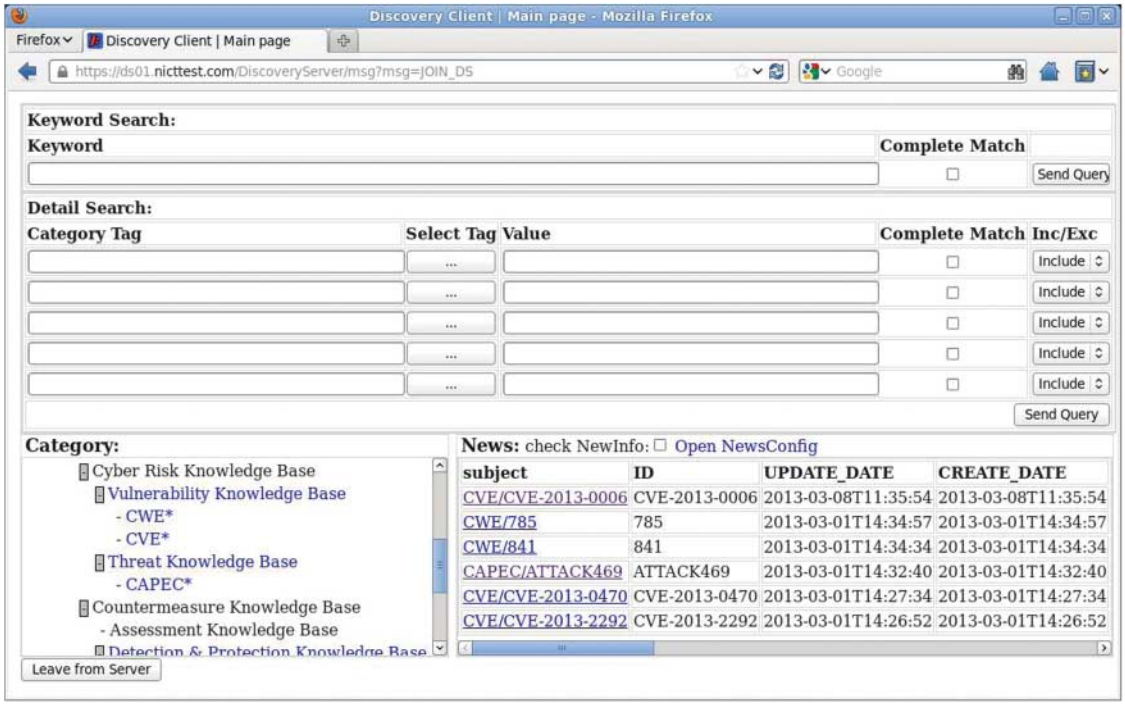

FIGURE 4. User interface for retrieving information.

ontology and knowledge base. Different types of information exist in different places. For instance, assorted cybersecurity information is publicly available in different repositories such as NVD, OSVDB and JVN. Moreover, there is confidential information that is stored inside an organization's private repository. In case of critical infrastructure, information on the industry control systems is also needed to protect the infrastructure, and such information is often stored separately from the cybersecurity information, be it public or private. The ontology makes it possible to have centralized administrative control over these kinds of information by building the knowledge base introduced in Section 6.3, although proper security measures including access control need to be implemented for the knowledge base.

Incident handling operations within an organization could also be streamlined. Current operations need roughly 1–3 h to realize and identify the need for security actions due to the time needed to confirm the incident information alerted by a detection system. Operators usually need to analyze IT assets and collect evidence of the incident; they analyze Event Records and produce Incident Records with which they evaluate the need to take countermeasures. One major reason for the long period of time required is that information is usually not well structured and is not easy to collect and compare. Once the information becomes well structured, the time needed for the above operations will be drastically reduced.

Moreover, detection, analysis and coordination of large-scale incidents could be streamlined. The structured information inside individual organizations could be converted into linked data [74] so that we can link the data beyond organizational borders. There are assorted issues to be dealt with to achieve this linkage, e.g. privacy of cybersecurity data, but we can facilitate and semi-automate the detection and analysis of large-scale incidents once the linkage is realized. Software can monitor event data inside multiple organizations, detect the occurrence of anomalies in different organizations at almost the same time, recognize the similarity of the abnormal events, then identify the occurrence of large-scale incidents. This facilitates the operations of SOCs and coordination centers.

We aim to realize these usage scenarios with the ontology in our future work. Through this work, we hope to reinforce cybersecurity and cyberdefense.

7. CONCLUSION

This paper proposed a *reference ontology* for cybersecurity operational information. Unlike other ontologies, this one was developed in close collaboration with cybersecurity organizations, including SOCs working in the USA, Japan and South Korea. It defines types of cybersecurity information along with the roles and operation domains, and clarifies who uses what types of information for what purposes.

This paper also demonstrated the ontology's usability. We used the ontology to review industry specifications; the ontology classified and organized industry specifications and demonstrated the applicability, coverage and effectiveness of current industry specifications. We also used the ontology to define an extensible information structure that orchestrates and collaborates with industry specifications; the information structure separates categories and formats of information, uses the ontology and industry specifications for the categories and formats, respectively, and links them. We then introduced a prototype of the cybersecurity information knowledge base that uses the ontology-based information structure. The knowledge base was able to handle assorted schemata and retrieve various kinds of information. Finally, we discussed streamlining cybersecurity operations with the ontology; the ontology will contribute to more efficient information management operations and eventually cybersecurity operations.

We believe this work contributes to the advancement of global cybersecurity information exchanges and the streamlining of cybersecurity operations. Nevertheless, the ontology should be regarded as the basis of such exchanges; further work is needed to encourage and expedite it. For instance, the cybersecurity knowledge base, introduced in Section 6.3, needs to be reinforced and operated online, so that it can become a publicly available online knowledge base. Non-technical issues such as motivation and privacy law issues regarding cybersecurity information exchanges also need to be addressed. We will continue working in this area in order to advance cybersecurity and its operations.

ACKNOWLEDGEMENTS

We would like to thank Inette Furey (DHS), Seon Meyong Heo (LAC Inc.), Robert Martin (MITRE Corporation), Kathleen Moriarty (EMC), Damir Rajnovic (Cisco Systems Ltd.), Anthony Rutkowski (Yaana Technologies), Gregg Schudel (Cisco Systems Ltd), Hiroshi Takechi (LAC Inc.) and Toshifumi Tokuda (IBM Inc.) for their many helpful comments, and their insightful perusal of our first draft.

FUNDING

This work was supported by JSPS KAKENHI [grant number 24700083]. Funding to pay the Open Access publication charges for this article was provided by JSPS KAKENHI [grant number 24700083].

REFERENCES

[1] Symantec Corporation (2014) *Internet Security Threat Report 2013*. Symantec Corporation, California, USA.

[2] IBM X-Force (2013) *IBM X-Force 2013 Mid-Year Trend and Risk Report*. IBM Corporation, New York, USA.

[3] OECD (2012) Cybersecurity Policy Making at a Turning Point: Analysing a New Generation of National Cybersecurity Strategies for the Internet Economy. *OECD Digital Economy Papers*, vol. 211, Paris, France.

[4] Kesh, S. and Ratnasingam, P. (2007) A knowledge architecture for IT security. *Commun. ACM*, **50**, 103–108.

[5] Martin, R.A. (2009) Making Security Measurable and Manageable. *CrossTalk, J. Def. Softw. Eng.*, **22**, 26–32.

[6] The MITRE Corporation (2014) *Making Security Measurable*. http://msm.mitre.org/.

[7] Gruber, T.R. (1995) Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, **43**, 907–928.

[8] Decker, S., Erdmann, M., Fensel, D. and Studer, R. (1999) Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. *Proc. IFIP TC2/WG2.6 8th Working Conf. Database Semantics—Semantic Issues in Multimedia Systems*, Rotorua, New Zealand, January 4–8, pp. 351–369. Kluwer Academic Publishers, Deventer, Netherlands.

[9] Tsoumas, B. and Gritzalis, D. (2006) Towards an Ontology-Based Security Management. *Proc. 20th Int. Conf. Advanced Information Networking and Applications*, Vienna, Austria, April 18–20, pp. 985–992. IEEE, New York, USA.

[10] Denker, G., Kagal, L. and Finin, T. (2005) Security in the semantic web using OWL. *Inf. Secur. Tech. Rep.*, **10**, 51–58.

[11] Fenz, S. and Ekelhart, A. (2009) Formalizing Information Security Knowledge. *Proc. 4th Int. Symp. Information, Computer, and Communications Security*, Sydney, Australia, March 10–12, pp. 183–194. ACM, New York, USA.

[12] Parkin, S.E., van Moorsel, A. and Coles, R. (2009) An Information Security Ontology Incorporating Human-Behavioural Implications. *Proc. 2nd Int. Conf. Security of Information and Networks*, Famagusta, North Cyprus, October 6–10, pp. 46–55. ACM, New York, USA.

[13] Wang, J.A. and Guo, M. (2009) OVM: An Ontology for Vulnerability Management. *Proc. 5th Annual Workshop on Cyber Security and Information Intelligence Research*, Tennessee, USA, January 8–10, pp. 1–4. ACM, New York, USA.

[14] Wang, J.A., Guo, M., Wang, H. and Zhou, L. (2012) Measuring and ranking attacks based on vulnerability analysis. *Inf. Syst. e-Bus. Manage.*, **10**, 455–490.

[15] NIST Interagency Report 7694 (2011) *Specification for the Asset Reporting Format 1.1*. National Institute of Standards and Technology, Maryland, USA.

[16] ITU-T X.1544 (2013) *Common Attack Pattern Enumeration and Classification*. International Telecommunications Union, Geneva, Switzerland.

[17] National Institute of Standards and Technology (2014) *Common Configuration Enumeration (CCE)*. http://nvd.nist.gov/cce/index.cfm.

[18] NIST Interagency Report 7502 (2010) *The Common Configuration Scoring System (CCSS): Metrics for Software Security Configuration Vulnerabilities*. National Institute of Standard and Technology, Maryland, USA.

[19] The MITRE Corporation (2014) *Common Event Expression*. http://cee.mitre.org/.

[20] NIST Interagency Report 7695 (2011) *Common Platform Enumeration: Naming Specification Version 2.3*. National Institute of Standards and Technology, Maryland, USA.
[21] The MITRE Corporation (2014) *Common Result Format Specification Version 0.3*. http://crf.mitre.org/.
[22] ITU-T X.1520 (2014) *Common Vulnerabilities and Exposures*. International Telecommunications Union, Geneva, Switzerland.
[23] Industry Consortium for Advancement of Security on the Internet (2014) *The Common Vulnerability Reporting Framework v1.1*. http://www.icasi.org/cvrf-1.1.
[24] ITU-T X.1521 (2011) *Common Vulnerability Scoring System*. International Telecommunications Union, Geneva, Switzerland.
[25] ITU-T X.1524 (2012) *Common Weakness Enumeration*. International Telecommunications Union, Geneva, Switzerland.
[26] The MITRE Corporation (2014) *Common Weakness Scoring System*. http://cwe.mitre.org/cwss/.
[27] The MITRE Corporation (2014) *Cyber Observable eXpression*. http://cybox.mitre.org/.
[28] Danyliw, R., Meijer, J. and Demchenko, Y. (2007) The Incident Object Description Exchange Format. *Request for Comments 5070*. Internet Engineering Task Force.
[29] ITU-T X.1546 (2014) *Malware Attribute Enumeration and Characterization*. International Telecommunications Union, Geneva, Switzerland.
[30] IEEE ICSG Malware Metadata Exchange Format Working Group (2014) *Malware Metadata Exchange Format Version 1.2*. http://grouper.ieee.org/groups/malware/malwg/Schema1.2/.
[31] NIST Interagency Report 7692 (2011) *Specification for the Open Checklist Interactive Language (OCIL) Version 2.0*. National Institute of Standards and Technology, Maryland, USA.
[32] ITU-T X.1526 (2014) *Language for the Open Definition of Vulnerabilities and for the Assessment of a System State*. International Telecommunications Union, Geneva, Switzerland.
[33] ISO/IEC 19770-2:2009 (2009) *Software Asset Management—Part 2: Software Identification Tag*. International Organization for Standardization/International Electrotechnical Commission, Geneva, Switzerland.
[34] GFD-R.192 (2011) *Web Services Agreement Specification (WS-Agreement)*. Open Grid Forum. Indiana, USA.
[35] xacml-3.0-core-spec-os-en (2013) *eXtensible Access Control Markup Language (XACML) Version 2.0*. Organization for the Advancement of Structured Information Standards, Massachusetts, USA.
[36] ISO/IEC 18180:2013 (2013) *Specification for the Extensible Configuration Checklist Description Format (XCCDF) Version 1.2*. International Organization for Standardization/International Electrotechnical Commission, Geneva, Switzerland.
[37] Lin, R.-R., Chang, Y.-H. and Chao, K.-M. (2011) Improving the performance of identifying contributors for xml keyword search. *ACM SIGMOD Rec.*, **40**, 5–10.
[38] Liu, Z., Walker, J. and Chen, Y. (2007) Xseek: A Semantic XML Search Engine Using Keywords. *Proc. 33rd Int. Conf. Very Large Data Bases*, Vienna, Austria, September 23–27, pp. 1330–1333. VLDB Endowment.
[39] Guo, L., Shao, F., Botev, C. and Shanmugasundaram, J. (2003) Xrank: Ranked Keyword Search Over XML Documents. *Proc. 2003 ACM SIGMOD Int. Conf. Management of Data*, California, USA, June 9–12, pp. 16–27. ACM, New York, USA.
[40] Baader, F., Horrocks, I. and Sattler, U. (2005) Description Logics as Ontology Languages for the Semantic Web. *Mechanizing Mathematical Reasoning*, Lecture Notes in Computer Science, pp. 228–248. Springer, Berlin, Germany.
[41] DSP0004 (2012) *Common Information Model (CIM) Infrastructure*. Distributed Management Task Force, Inc., Oregon, USA.
[42] REC-owl2-overview-20121211 (2012) *OWL 2 Web Ontology Language Document Overview (Second Edition)*. The World Wide Web Consortium.
[43] Masoumzadeh, A. and Joshi, J. (2013) Privacy Settings in Social Networking Systems: What You Cannot Control. *Proc. 8th ACM SIGSAC Symp. Information, Computer and Communications Security*, Hangzhou, China, May 8–10, pp. 149–154. ACM, New York, USA.
[44] Blanco, C., Lasheras, J., Fernández-Medina, E., Valencia-García, R. and Toval, A. (2011) Basis for an integrated security ontology according to a systematic review of existing proposals. *Comput. Stand. Interfaces*, **33**, 372–388.
[45] ISO/IEC 27032:2012 (2012) *Guidelines for Cybersecurity*. International Organization for Standardization/International Electrotechnical Commission, Geneva, Switzerland.
[46] ITU-T E.409 (2004) *Incident Organization and Security Incident Handling: Guidelines for Telecommunication Organizations*. International Telecommunications Union, Geneva, Switzerland.
[47] ITU-T X.1500 (2011) *Overview of Cybersecurity Information Exchange (CYBEX)*. International Telecommunications Union, Geneva, Switzerland.
[48] Rutkowski, A. *et al.* (2010) Cybex—the cybersecurity information exchange framework (x.1500). *Comput. Commun. Rev.*, **40**, 59–64.
[49] Brownlee, N. and Guttman, E. (1998) Expectations for Computer Security Incident Response. *Request for Comments 2350*. Internet Engineering Task Force.
[50] National Institute of Standards and Technology (2014) *Special Publications (800 Series)*. http://csrc.nist.gov/publications/PubsSPs.html.
[51] Special Publication 800-12 (1995) *An Introduction to Computer Security: The NIST handbook*. National Institute of Standards and Technology, Maryland, USA.
[52] Special Publication 800-35 (2003) *Guide to Information Technology Security Services*. National Institute of Standards and Technology, Maryland, USA.
[53] Special Publication 800-61 Revision 2 (2012) *Computer Security Incident Handling Guide*. National Institute of Standards and Technology, Maryland, USA.
[54] Special Publication 800-86 (2006) *Guide to Integrating Forensic Techniques into Incident Response*. National Institute of Standards and Technology, Maryland, USA.
[55] Special Publication 800-115 (2008) *Technical Guide to Information Security Testing and Assessment*. National Institute of Standards and Technology, Maryland, USA.
[56] Special Publication 800-55 (2008) *Performance Measurement Guide for Information Security*. National Institute of Standards and Technology, Maryland, USA.
[57] COBIT 5 (2012) A Business Framework for the Governance and Management of Enterprise IT. *ISACA*, Illinois, USA.

[58] National Institute of Standards and Technology (2014) *Framework for Improving Critical Infrastructure Cybersecurity*. National Institute of Standards and Technology, Maryland, USA.

[59] The White House (2013) Executive Order 13636—Improving Critical Infrastructure Cybersecurity. *Fed. Register*, **78**.

[60] Schmitt, M.N. (2013) *Tallinn Manual on the International Law Applicable to Cyber Warfare*. Cambridge University Press, Cambridge, UK.

[61] National Institute of Standards and Technology (2014) *National Vulnerability Database Version 2.2*. http://nvd.nist.gov/.

[62] Open Security Foundation (2014) *Open Sourced Vulnerability Database*. http://osvdb.org/.

[63] Red Hat Inc. (2014) *Security Measurement*. https://www.redhat.com/security/data/metrics/.

[64] JPCERT/CC and IPA (2014) *Japan Vulnerability Notes*. http://jvn.jp/.

[65] REC-rdf-concepts-20040210 (2004) *Resource Description Framework (RDF): Concepts and Abstract Syntax*. The World Wide Web Consortium.

[66] C(92)188/FINAL (1992) *OECD Guidelines for the Security of Information Systems*. Organisation for Economic Co-operation and Development, Paris, France.

[67] ISO/IEC 13335-1:1996 (2004) *Management of Information and Communications Technology Security—Part 1: Concepts and Models for Information and Communications Technology Security Management*. International Organization for Standardization/International Electrotechnical Commission, Geneva, Switzerland.

[68] Moreau, L. *et al.* (2008) The provenance of electronic data. *Commun. ACM*, **51**, 52–58.

[69] Johansson, P., Hall, L., Sikstrom, S. and Olsson, A. (2005) Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, **310**, 116.

[70] Stanford Center for Biomedical Informatics Research (2014) *The Protégé Ontology Editor and Knowledge Acquisition System*. http://protege.stanford.edu/.

[71] Danyliw, R. and Stoecker, P. (2014) The Incident Object Description Exchange Format v2. *Internet-Draft draft-ietf-mile-rfc5070-bis-06*. Internet Engineering Task Force.

[72] Apache Jena (2014) *SDB—Persistent Triple Stores Using Relational Databases*. http://jena.apache.org/documentation/sdb/index.html.

[73] REC-sparql11-overview-20130321 (2011) *SPARQL Query Language for RDF*. The World Wide Web Consortium.

[74] Bizer, C., Heath, T. and Berners-Lee, T. (2009) Linked Data—The Story So Far. *Int. J. Semant. Web Inf. Syst.*, **5**, 1–22.