Representing and Reasoning About Logical Network Topologies

Shaun Voigt, Catherine Howard(✉), Dean Philp,
and Christopher Penny

Defence Science and Technology Group, Edinburgh, Adelaide, Australia
{shaun.voigt, catherine.howard,
dean.philp}@dst.defence.gov.au

Abstract. For network analysts, constructing a representation, and developing an understanding, of logical network topologies is crucial for a wide range of cyber security applications. However, constructing a representation of logical network topologies is difficult. This paper presents three novel ontologies; the Internet Protocol (IP) Ontology, the Open Shortest Path First (OSPF) Ontology and the Border Gateway Protocol (BGP) Ontology. These ontologies provide a common, technology independent syntax and semantics for complex communication network concepts. The semantic and syntactic interoperability provided by these ontologies enables data from disparate, heterogeneous sources, such as network diagrams, router configuration files and routing protocol messages, to be consistently represented, which facilitates information fusion. The approach presented in this paper allows domain knowledge to be encoded in an intuitive manner, facilitates knowledge discovery by automated reasoning, and facilitates the process of making specialist knowledge and tradecraft accessible to non-expert network analysts.

Keywords: Ontologies · Network data · Network topologies

1 Introduction

For network analysts, constructing a representation, and developing an understanding, of logical network topologies[1] is crucial for a wide range of cyber security applications such as traffic path estimation, network monitoring and management [1], network vulnerability assessment and defence [2], identifying network boundaries and understanding the propagation of BGP hijacks. However, constructing a representation of logical network topologies is difficult, especially at Internet scale. The Internet is the largest, most complex artificially deployed system in existence [3] and there are many disparate, heterogeneous sources of data which could potentially be used. The application of automated information fusion techniques, and the associated underlying

[1] The topology of a network is the arrangement of the various network elements, such as routers, computers and links, within the network. The topology of a network may be depicted physically or logically. The physical topology of a network is the arrangement of the physical components of the network, including the location of devices and cables. While the logical topology illustrates how information flows through the network.

M. Croitoru et al. (Eds.): GKR 2017, LNAI 10775, pp. 73–83, 2018.
https://doi.org/10.1007/978-3-319-78102-0_4

knowledge representation and automated reasoning techniques, could assist in addressing these scale and complexity issues. The fusion of data from multiple sources can provide a more detailed representation of the logical network topology than the representation provided by any individual data source in isolation. However, the automated fusion of data from disparate, heterogeneous sources requires semantic and syntactic interoperability. To provide this interoperability, this research adopted an ontological approach to knowledge representation.

There is a dearth of literature on the use of ontologies for constructing representations of logical communication network topologies; ontologies have been developed for network planning and design (e.g., [4]), network measurement and monitoring (e.g., [5]) and the provisioning, configuration and management of virtual or physical network resources (e.g., [6–13]). However, none of these ontologies provided a formal specification of the IP, OSPF and BGP concepts required by this research, at the required level of detail. Hence this research developed three novel ontologies which can be used to represent complex communication network concepts; the Internet Protocol (IP) Ontology, the Open Shortest Path First (OSPF) Ontology and the Border Gateway Protocol (BGP) Ontology. This paper presents these three novel ontologies.

The rest of this paper is structured as follows. Section 2 describes the data sources utilised by this research. Section 3 presents the three novel ontologies, justifies the selection of the Web Ontology Language (OWL) as the implementation language and describes the knowledge representation and reasoning processes. Section 4 provides an example of producing a representation of a logical network topology using some of the data sources described in Sect. 2 and the ontologies and processes outlined in Sect. 3. Section 5 presents a brief discussion while Sect. 6 presents the conclusions.

2 The Data Sources

There are many disparate, heterogeneous sources of network data which could potentially be used to construct representations of logical network topologies. This research focused on being able to represent, fuse and reason about six such sources; network diagrams, router configuration files, routing tables, Open Shortest Path First (OSPF) Link State Advertisements (LSAs), Border Gateway Protocol (BGP) update messages and open source data.

A network diagram is a visual representation of the physical or logical topology of a network. It depicts the nodes (including routers, switches, servers, printers and hosts) in the network and the connections between them.

A router's configuration file contains all the commands required to configure the router. It contains information such as the IP addresses of the router's interfaces, the routing protocols used on each interface and the metrics used by link state routing protocols[2].

For each reachable destination, a routing table lists the network element which is next along the path to the destination. When an IP packet arrives, a router uses this

[2] In a link state routing protocol, each router constructs a map of the connectivity of the network in which it resides.

table to determine the interface on which to forward the packet based on its destination IP address.

OSPF [14] is the most widely used interior gateway protocol[3] (IGP) on the Internet [15]. Link State Advertisements (LSAs) are the basic communication mechanism of OSPF. There are eleven different types of LSAs. This research utilises Router (also referred to as Type 1) and Network (also referred to as Type 2) LSAs. A Router LSA contains information about all routers and networks which are directly connected to the originating router. A Network LSA includes the network identifier, subnet mask and a list of routers which are joined together by the broadcast domain[4].

BGP [16] is an exterior gateway protocol; it is used to facilitate inter Autonomous System[5] (AS) relationships by exchanging routing and reachability information among ASes on the Internet. When a BGP session is initialised between routers, update messages are sent to exchange routing information until the complete BGP routing table has been exchanged. A router advertises the networks which are reachable via each of its neighbours and how many hops away each network is.

There is a myriad of open source information which could potentially be useful. This research focused on utilising some of the data available from the Center for Applied Internet Data Analysis (CAIDA)[6] [17], including the:

- AS name, number and owner;
- Networks that an AS is the registered owner of;
- Networks advertised by an AS; and
- Inter-AS relationships that an AS participates in (i.e., peering and customer-provider relationships).

From the above descriptions, it can be seen that the six data types are disparate and heterogeneous. The data itself is complex, relational data, which is not easily understood by analysts without specialist communication network knowledge and experience.

3 The Ontologies and the Knowledge Representation and Reasoning Process

The Web Ontology Language (OWL) [18] was selected to implement the three ontologies because, among other reasons:

- OWL is explicitly designed to support the integration of data from multiple sources [19].

[3] Interior gateway protocols manage the routing of traffic within individual ASes.

[4] A broadcast domain is a logical division of a network, in which all devices can reach each other by broadcast at the data link layer. For example, a multi-access network is a single broadcast domain. Ethernet is also an example of a broadcast domain.

[5] An Autonomous System is a network, or collection of networks, which are managed or supervised by a single administrative entity or organisation.

[6] CAIDA is a collaboration of government, research and commercial entities aimed at promoting greater cooperation in the engineering and maintenance of the global Internet infrastructure.

- OWL and RDF are well suited to the representation of complex, relational data [19].
- RDF triples can be represented as semantic networks [20], which are a natural representational match for logical network topologies (which can be represented as undirected graphs [19]).
- OWL provides an explicit separation between syntax and semantics.
- OWL can be coupled with semantic reasoners and rule-based languages, such as the Semantic Web Rule Language (SWRL) [18], to support automated reasoning.
- OWL allows ontologies to reuse classes and properties from existing, published ontologies [12, 19].

In this research, OWL and RDF were used during the knowledge representation process and SWRL and the SPARQL Protocol and RDF Query Language (SPARQL) [18] were used during the reasoning process.

Figure 1 shows the hierarchy of the IP, OSPF and BGP ontologies, with the OSPF and BGP ontologies inheriting classes, data properties and object properties from the IP ontology. Because there is insufficient room to present the full OWL functional syntax, the ontologies will be presented using relational diagrams. Relational diagrams depict the set of classes, data properties and object properties in an ontology. In relational diagrams, classes are represented by large rectangles, with the name of the class in bold print, and object properties are underlined and linked to their range types by directed lines.

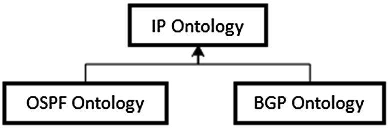

Fig. 1. The inheritance hierarchy of the ontologies.

The IP Ontology, shown in Fig. 2, represents concepts at the IP layer (i.e., Layer 3 of the OSI model [21]). As the IP ontology focuses on the IP layer, Layer 2 devices (such as switches) and physical connections (such as cables) are not included. The set of classes in the IP ontology is **C** = {*Network Element*, *Network*, *Router*, *Computer, Interface*, *Route Entry*, *Default Route Entry*, *Directly Connected Route Entry*}.

The OSPF ontology, shown in Fig. 3, extends the IP Ontology by introducing OSPF specific concepts such as OSPF areas[7] and Area Border Routers (ABRs)[8]. The set of classes in the ontology is **C** = {*Network Element*, *Network*, *Router*, *Computer, Interface*, *Area, Route Entry*, *Default Route Entry*, *Directly Connected Route Entry, OSPF Summary Route Entry*}. From Fig. 3 it can be seen that the OSPF Ontology inherits classes from the IP ontology (e.g., the *Network Element*, *Network* and *Router* classes), specialises some of the object properties of these classes (e.g., the *hasNeighbour* object property of the *Router* class has a new *hasOSPFRouterNeighbour*

[7] An OSPF network can be subdivided into multiple routing areas in order to simplify administration or optimise traffic flow or resource utilisation.

[8] ABRs are routers which have interfaces in multiple areas.

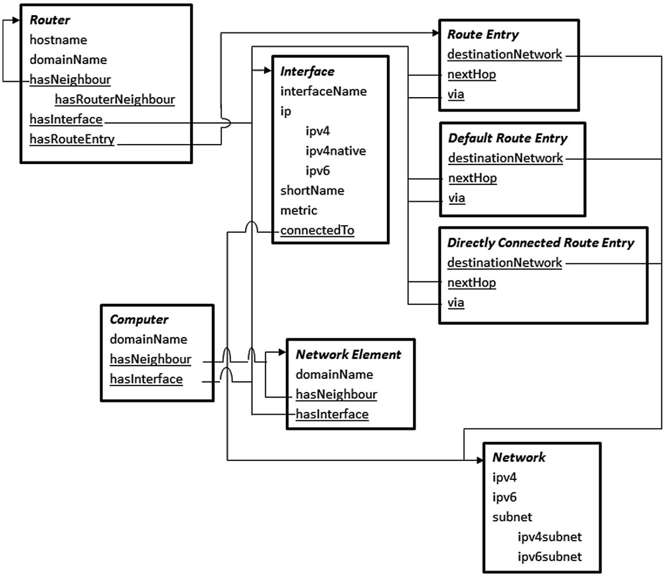

Fig. 2. The relational diagram of the IP ontology.

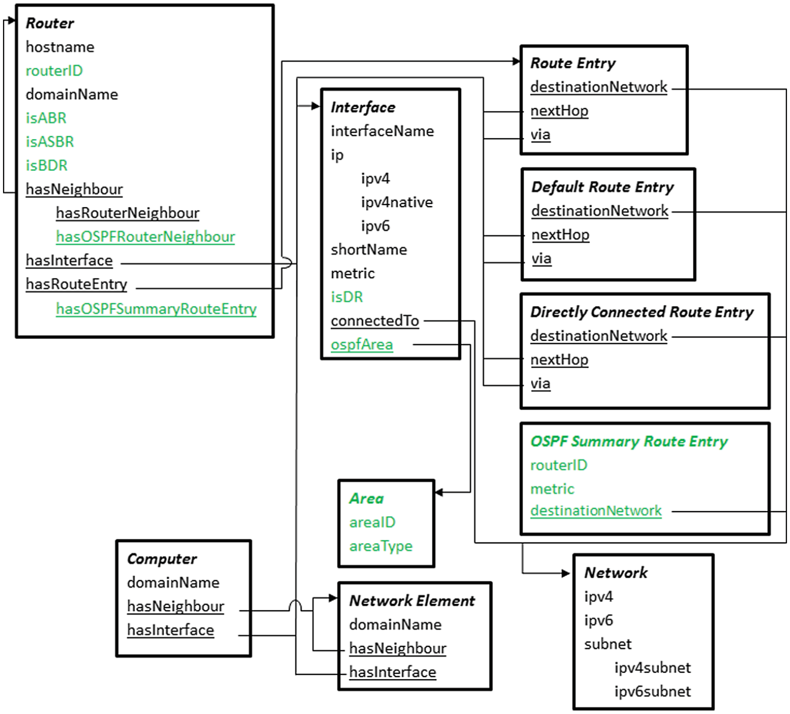

Fig. 3. The relational diagram of the OSPF ontology. (Color figure online)

specialisation), adds new data properties (e.g., the *isABR*, *isASBR* and *isBDR* properties of the *Router* class) and adds new classes such as the *Area* and *OSPF Summary Route Entry* classes.

The BGP ontology, shown in Fig. 4, extends the IP Ontology by introducing BGP specific concepts, such as update messages, ASes and AS paths. The set of classes in the ontology is **C** = {*Network Element*, *Network*, *Router*, *Interface*, *Route Entry*, *Autonomous System, Update Message, AS Path}*. From Fig. 4 it can be seen that the BGP Ontology inherits the classes from the IP ontology (e.g., the *Network Element*, *Network* and *Router* classes), adds new object properties (e.g., the *eBGPNeighbour* and *iBGPNeighbour* properties of the *Interface* class) and adds new classes such as the *Update Message*, *Autonomous System* and *AS Path* classes. In Figs. 3 and 4, the classes and properties inherited from the IP Ontology are shown in black, while the new or specialised classes and properties are shown in green.

During the knowledge representation process, the IP, OSPF and BGP ontologies provide a common, technology independent[9] syntax and semantics for complex communication network concepts, so that heterogeneous data can be encoded into a

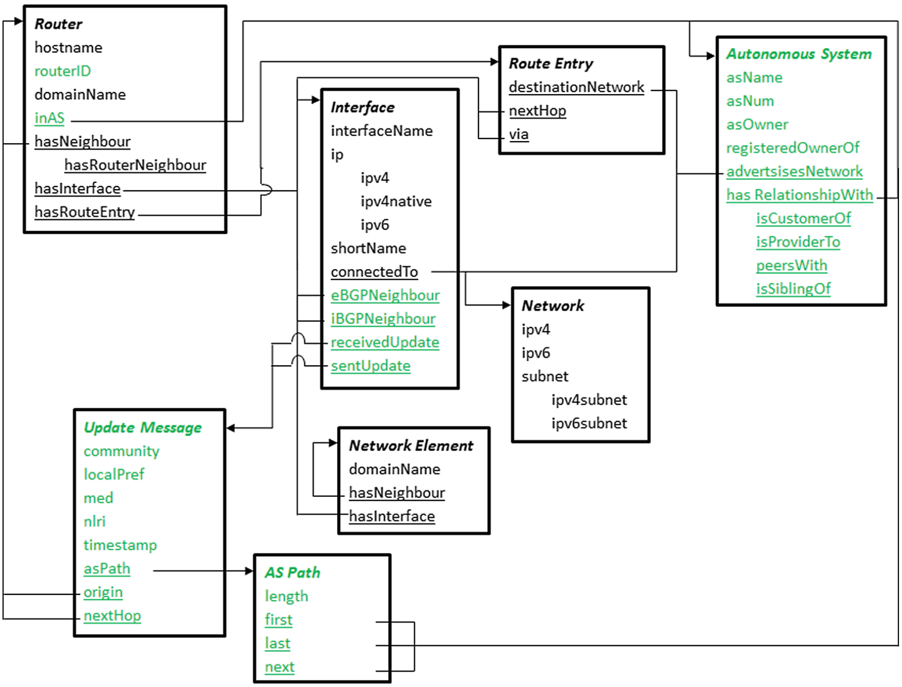

Fig. 4. The relational diagram of the BGP ontology. (Color figure online)

[9] For example, as a result of slight differences in their interpretation of the Internet Engineering Task Force (IETF) OSPF standards, Cisco and Juniper routers implement OSPF in different ways. The OSPF ontology presented in this section, however, provides a generic representation of OSPF which is not dependant on the specific implementation technology.

consistent representation. Once encoded, the data is in RDF triple format and is referred to as instance data. The instance data are stored in a triple store in the knowledge base.

Context specific rules enable subject matter experts (SMEs) to encode specialist knowledge or tradecraft using SWRL. Examples of context specific rules are provided in Sect. 4. Context specific rules can be used to perform data cleaning and information fusion. During the reasoning process, using the ontologies and context specific rules, the rule-based inference engine performs reasoning over the instance data in the knowledge base. The reasoning process is a forward-chaining, data-driven process, whereby new information can trigger the execution of additional context specific rules.

4 Fusion Example

Consider the scenario shown in Fig. 5. In this scenario, there are two ASes, *AS10143* and *AS1221*, which are connected by a single inter-AS relationship. *AS10143* has two routers *AS10143R1* and *AS10143R2*. *AS1221* has one router *AS1221R1*. *AS1221R1* has an external BGP relationship with *AS10143R1*. *AS10143* is using OSPF as its IGP. Suppose that the available information sources include:

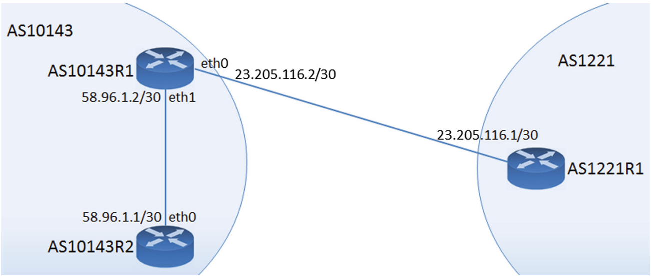

Fig. 5. The scenario under consideration.

- Open source CAIDA data pertaining to *AS10143* and *AS1221*;
- A BGP update message sent from *AS10143R1* to *AS1221R1;*
- Router configuration files for *AS10143R1* and *AS10143R2;* and
- OSPF Router LSAs issued by *AS10143R1* and *AS10143R2*.

However, for this example, it is assumed that no router configuration files, OSPF Router LSAs or BGP update messages are available for *AS1221*.

Using context specific rules such as:

$$\begin{aligned}&\textit{If two Network objects have the same ipv4subnet value},\\&\textit{then the two Networks objects are the same object};\end{aligned} \tag{1}$$

$$\begin{aligned}&\textit{If two Interface objects have the same ipv4 value},\\&\textit{then the two Interface objects are the same object};\end{aligned} \tag{2}$$

$$\begin{aligned}&\textit{If two Router objects have the same routerID value},\\ &\textit{then the two Router objects are the same object}; \text{ and}\end{aligned} \tag{3}$$

$$\begin{aligned}&\textit{If two AS objects have the same asNum},\ \ \textit{and it is a public asNum},\\ &\textit{then it is the same AS},\end{aligned} \tag{4}$$

the data from the aforementioned sources can be fused to produce the representation of the logical network topology shown in Fig. 6. This semantic network contains

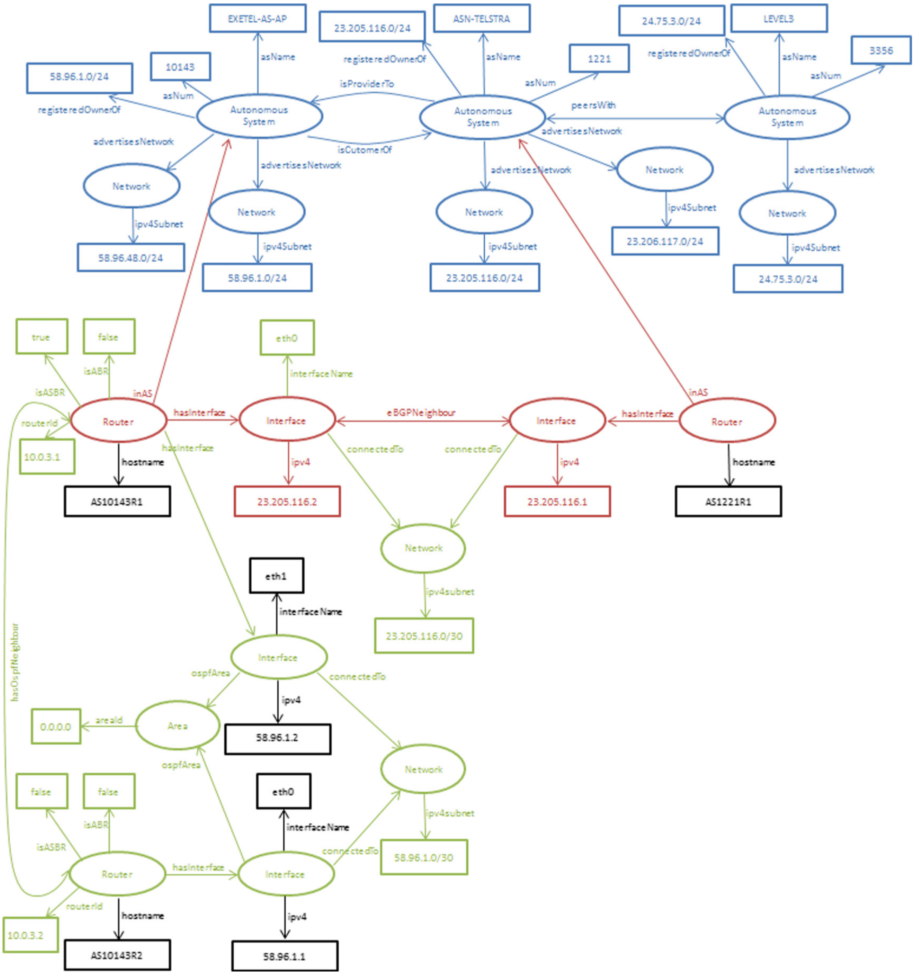

Fig. 6. The semantic network resulting from the fusion of all the available data. Blue represents open source CAIDA data, red represents the data obtained from the BGP update message sent from *AS10143R1* to *AS1221R1*, black represents the data obtained from *AS10143R1*'s and *AS10143R2*'s configuration files and green represents the data obtained from *AS10143R1*'s and *AS10143R2*'s LSAs. (Color figure online)

information about interfaces, routers, networks, areas and ASes and the relationships between these concepts. It combines intra-AS connectivity information provided by OSPF with inter-AS connectivity information provided by BGP and CAIDA, allowing an analyst to see the connection between Internet level routing and the private network infrastructure of EXETEL-AS-AP[10]. It can be seen that the enriched representation of the network's logical topology provided by Fig. 6 is more detailed and accurate than the representation provided by any individual data source in isolation.

5 Discussion

The IP, OSPF and BGP ontologies provide a consistent way to represent complex communication network concepts. The ontologies are easily extensible. They support communication and information sharing, automated reasoning and the reuse of domain knowledge. They also limit ambiguity and make domain assumptions explicit. Making the domain assumptions explicit makes it possible to change these assumptions if the knowledge about the domain changes. Being able to represent the resulting network topologies as semantic networks facilitates human understanding.

The three ontologies all contain concepts at a range of abstractions; from high level concepts such as ASes and networks through to low level concepts such as router interfaces. This allows:

- Information at different levels of abstraction to be represented and fused. For example, Fig. 6 depicts EXETEL-AS-AP in a high level of detail while ASN-TELSTRA, where less information is available, is represented at a more abstract level.
- Networks to be represented at different levels of abstraction. For example, using the same ontologies and same data sources, the same network could be represented by a semantic network containing:
  - Networks, routers, interfaces, IP addresses, subnets, interface names, host names and the *hasInterface* and *connectedTo* relationships; or
  - Routers, host names and the *hasRouterNeighbour* relationships.
- The same concepts to be used in different ways.

Being able to represent information at different levels of abstraction is important because abstraction allows:

- Complex data to be simplified. This simplification can reduce the complexity of both data analysis and visualisation and can enable complex data to be hidden from non-expert analysts.
- The application of graph theoretic techniques at an abstracted level, rather than the lowest level of detail, where the size and complexity of the semantic network may preclude their use.

[10] Synthetic data has been used for the private network infrastructure in order to demonstrate the fusion techniques.

The context specific rules discussed in Sect. 3 can provide a natural way for SMEs to encode their specialist knowledge or tradecraft, potentially making this knowledge more accessible to non-experts analysts. Because OWL is a declarative language, rules can be developed which work for a large number of instances. Rules can be generic, and hence applicable to all types of networks, or they can be specific for specific types of networks (for example, content distribution networks). The ability to use different context specific rule sets, based on the situation, provides a level of flexibility. However rules can have a number of limitations. For example, the quality of the rule base developed for a particular domain will be dependent on the experience and point of view of the SMEs who construct it, so there may be gaps, overlaps and inconsistencies. Encoding rules can also be difficult; knowledge elicitation is manual and can be error prone. It can be easy, for example, to create contradictory rules. A large rule set can be difficult to maintain and update.

6 Conclusions

This paper presented three novel ontologies; the IP Ontology, the OSPF Ontology and the BGP Ontology. These ontologies provide a common, technology independent syntax and semantics for complex communication network concepts. The semantic and syntactic interoperability provided by the three ontologies allows data from disparate, heterogeneous sources to be consistently represented, which facilitates information fusion.

The approach presented in this document allows domain knowledge to be encoded in an intuitive manner, facilitates knowledge discovery by automated reasoning, and facilitates the process of making specialist knowledge and tradecraft accessible to non-expert network analysts.

While ontological approaches to knowledge representation have many strengths, the quality of an ontology developed for a particular domain will always be dependent on the experience and point of view of the SMEs who build it, so there are always gaps, overlaps and inconsistencies. However, this is true of any knowledge representation technique.

Acknowledgements. Part of this work was conducted using the Protégé resource [22], which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

References

1. van der Ham, J., Ghijsen, M., Grosso, P., de Laat, C.: Trends in Computer Network Modeling Towards the Future Internet. https://arxiv.org/pdf/1402.3951v2.pdf. Accessed Oct 2016
2. Motamedi, R., Rejaie, R., Willinger, W.: A survey of techniques for internet topology discovery. IEEE Commun. Surv. Tutor. **17**(2), 1044–1065 (2013)
3. Ioannou, P.A., Pitsillides, A.: Modeling and Control of Complex Systems. CRC Press, Boca Raton (2008)

4. Rahman, M., Pakstas, A., Wang, F.Z.: Towards communications network modelling ontology for designers and researchers. In: Proceedings of the International Conference on Intelligent Engineering Systems, London, England (2006)
5. MOMENT - Monitoring and Measurement in the Next Generation Technologies. http://www.salzburgresearch.at/en/projekt/moment_en/. Accessed Oct 2016
6. Yeung, D., Qu, Y., Zhang, J., Chen, I., Lindem, A.: Yang Data Model for OSPF Protocol. https://tools.ietf.org/html/draft-ietf-ospf-yang-01. Accessed Oct 2016
7. Zhdankin, A., Patel, K., Clemm, A., Hares, S., Jethanandani, M., Liu, X.: Yang Data Model for BGP Protocol. https://tools.ietf.org/html/draft-zhdankin-idr-bgp-cfg-00. Accessed Oct 2016
8. Common Information Model. http://www.dmtf.org/standards/cim. Accessed Aug 2015
9. Strassner, J.: DEN-ng: achieving business-driven network management. In: Proceedings of the IEEE/IFIP Network Operations and Management Symposium (2002)
10. van der Ham, J., Dijkstra, F., Lapacz, R., Brown, A.: The network markup language; a standardized network topology abstraction for inter-domain and cross-layer network applications. In: Proceedings of the TERENA Networking Conference, Maastricht, Netherlands (2013)
11. van der Ham, J., Dijkstra, F., Travostino, F., Andree, H., de Laat, C.: Using RDF to describe networks. Future Gener. Comput. Syst. **22**(8), 862–867 (2006)
12. Ghijsen, M., van der Ham, J., Grosso, P., Dumitru, C., Zhu, H., Zhao, Z., de Laat, C.: A semantic-web approach for modelling computing infrastructures. J. Comput. Electr. Eng. **39**, 2553–2565 (2013)
13. Network Innovation over Virtualized Infrastructures. http://www.fp7-novi.eu/index.php. Accessed Oct 2016
14. Moy, J.: RFC 2328 - OSPF Version 2. https://www.ietf.org/rfc/rfc2328.txt. Accessed Oct 2016
15. Nakibly, G., Gonikman, D., Kirshon, A., Boneh, D.: Persistent OSPF attacks. In: Proceedings of the Nineteenth Annual Network and Distributed System Security Conference (2012)
16. Rekhter, Y., Li, T., Hares, S.: RFC 4271 - A Border Gateway Protocol 4 (BGP-4). https://www.ietf.org/rfc/rfc4271.txt. Accessed Oct 2016
17. Center for Applied Internet Data Analysis. www.caida.org. Accessed Oct 2016
18. Antoniou, G., van Harmelen, F.: A Semantic Web Primer. MIT Press, Cambridge (2004)
19. Reynolds, D., Thompson, C., Mukerji, J., Coleman, D.: An Assessment of RDF/OWL Modelling. Digital Media Systems Laboratory, HP Laboratories Bristol, HPL-2005-189 (2005)
20. Sowa, J.: Semantic networks. In: The Encyclopedia of Artificial Intelligence, 2nd edn. (1987)
21. OSI Model. https://en.wikipedia.org/wiki/OSI_model. Accessed Oct 2016
22. Protege. http://protege.standford.edu/. Accessed Oct 2016