ISRF: interest semantic reasoning based fog firewall for information-centric Internet of Vehicles

Yi Liu1 1, Jun Wu1, Jianhua Li1, Wu Yang2, Hao Chen1, Gaolei Li1
1School of Cyber Security, Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai 200240, People’s Republic of China
2Information Security Research Center, Harbin Engineering University, Harbin 150001, People’s Republic of China
E-mail: ly2017@sjtu.edu.cn

Abstract: Nowadays, Internet of Vehicles (IoV) has attracted lots of attention as an important component of intelligent transportation. Meanwhile, information-centric network is a meaningful next-generation network with content-oriented attributes, which has been introduced into IoV for more flexible services. In information-centric IoV, there are abundant semantic features because of the content big data in the network layer. As very few works of information-centric IoV consider semantic models, current security approaches in information-centric IoV cannot defend against efficiently novel threats with penetrated malicious content and semantic obfuscation, especially at the network edge. Thus an efficient firewall is a must for information-centric IoV. However, existing firewalls based on traditional computing resources cannot detect content threats implicated in semantics at the edge of networks. In this study, the authors propose an interest semantic reasoning based fog firewall for information-centric IoV. Firstly, a fog firewall is proposed to construct an edge isolation for information-centric IoV. Moreover, interest semantic reasoning with context weight and knowledge graph is proposed to mine implicit relations and illegal knowledge in names and contents. Based on knowledge constructed from reasoning, it aims to perceive threats from interest packet in information-centric IoV. Simulations verify its feasibility and efficiency of the proposed scheme.

1 Introduction

In the new era of Internet of Things (IoT), researchers focus on launching the development of connected vehicle networks into the Internet of Vehicles (IoV). Vehicles are connected to the network to communicate with other vehicles, roadside units (RSUs), infrastructure and cloud centres [1]. IoV aims to provide timely and reliable communication of hazard/warning and congestion information among vehicles to improve the safety and driving experiences [2]. In addition, vehicles are expected to act as personal computers provide entertainment and business functions. Therefore, plenty of IoV services have appeared. For example, traffic condition broadcasting services avoid traffic jams and danger predication and alert services provide safe driving experiences. All these services above aim to provide extensive information locally. Moreover, some conventional Internet applications also occupy a certain proportion of IoV services to provide better travelling experiences, such as social services, multimedia applications and so on. Numerous vehicles are equipped with on-board computers to provide entertainment and business services to passengers. These kinds of vehicles have some commons that they are big data streams and tend to be shareable. Thus, service-oriented information processing and transmission dominates the IoV environment [3].

However, the demand of exponentially growing IoV traffic is changing from IP-based communications to content-centric communications such as multimedia streaming and information sharing [2]. On the one hand, IP address-oriented connections are difficult to adapt high mobility, localised communications and extreme dynamic environments of IoV. The TCP/IP protocol needs to be maintained when high mobility is damaged the connection. On the other hand, the rapid changing network topology in a mobile environment tends to result in inefficient routing. Therefore, information-centric network (ICN) is regarded as an alternative to IP networks in IoV. ICN is a promising paradigm for the next generation networks [4]. It focuses on providing information-oriented network protocols, including content-centric subscribe mechanism and semantic-dominated naming, routing and caching strategies. Therefore, the unique information-centric IoV structure leads to abundant semantic features and appliance of content subscription and distribution as well as cache. The reasons why ICN is introduced into IoV are as follows. First, different from the IP end-to-end network architecture, ICN supports data transmission even in unreliable and extreme environments, like frequently dynamic connections for mobile and ubiquitous computation [5]. Second, in most occasions, users in IoV do not care about the sources of information or services, but the content itself. In addition, some popular data tend to be frequently requested in some conditions or locations, in this case, the in-network cache can provide low-latency services to terminal users. Therefore, many existing works have studied the information-centric IoV. And the scene diagram of information-centric IoV is shown in Fig. 1.

Featured with a content-driven subscription, information-centric IoV suffers from novel penetration threats that are implicated in content. Packet names are generated from content attributes and utilise human-readable languages, thus causing the potential of semantic vagueness and ambiguity. Firstly, names of interest packets could have potential semantic features and relevance with illegality. Moreover, attackers tend to send implicitly forbidden content that is obfuscating and correlated with requests in some semantic dimensions, which can lead to the failure of detection with the inherent firewall blacklists. Therefore, the content-oriented subscribe mechanism offers the potential for content obfuscation and spoofing attacks [6]. It is important to perceive content threats to identify various illegal features in content. Defence of penetrated content threats needs semantic reasoning analysis to identify the inherent and potential illegality features of content and to mine knowledge relations between traffic. Therefore, semantic based security knowledge policies are required on top of the traditional defence. However, existing works of information-centric IoV security focus on attacks of naming, routing and caching dispersedly [7]. They neglect content semantic analysis to mine potential threats and construct knowledge relations through the nature of the different content. Moreover, lacking in content awareness and semantic analysis, existing works...
cannot implement potential threat awareness of packet content or customise defence policy for the content of each distinct packet.

Moreover, conventional TCP/IP firewalls are incompatible with information-centric IoV protocols and data structures. Firstly, content-oriented packet forms and subscription schemes of information-centric IoV limit the function migration of TCP/IP firewalls. Secondly, existing firewalls depend on traditional computation paradigm and fail to conduct content semantic analysis and content threat awareness on the edge of networks. Thirdly, blacklists and blocking policies of existing firewalls are created in the cloud centre uniformly, all the firewalls apply unified policies without any analysing tools and neglecting the heterogeneity of different network environments, which causes numerous policy errors. Last but not least, current firewall databases of threat information are deployed in the cloud centre, it is inconvenient to upload and share the various kinds of novel semantic threats in information-centric IoV dynamically. Thus, existing firewalls cannot construct the isolation defence system for information-centric IoV.

To address aforementioned challenges, we propose a fog-enabled firewall for information-centric IoV to provide isolation defence as well as content-threat awareness with interest semantic reasoning and policy customisation. To perform the edge isolation and semantic content awareness for information-centric IoV, fog computing, known as transferring intelligence and data from cloud centre to edge network [8], is adapted to offer computation and edge distribution in the proposed firewall system. We employ fog computing to support context and content awareness of information-centric IoV packets and conduct the semantic analysis and customised configuration of information-centric IoV firewalls. Additionally, a semantic reasoning approach based on knowledge graph is designed. It aims to construct security knowledge and mine possible potential relations between the requested content and blacklists involving the illegal information. Besides, to defend latent and continuous attacks, the proposed firewall model collects context traffic correlated with pending packets, which emphasises relevant semantic dimensions to guide the reasoning direction. Semantic reasoning analyses the content attributes through interest names and thus predicts potential threats of pending packets and their response data. Customised policies for each packet are generated as a result of semantic analysis. Moreover, customised configuration supports users to specify filtering policy to exclude unexpected responses beyond the blacklists. Our contributions are as follows:

Additionally, a semantic reasoning approach based on knowledge graph is designed. It aims to construct security knowledge and mine possible potential relations between the requested content and blacklists involving the illegal information. Besides, to defend latent and continuous attacks, the proposed firewall model collects context traffic correlated with pending packets, which emphasises relevant semantic dimensions to guide the reasoning direction. Semantic reasoning analyses the content attributes through interest names and thus predicts potential threats of pending packets and their response data. Customised policies for each packet are generated as a result of semantic analysis.

Moreover, customised configuration supports users to specify filtering policy to exclude unexpected responses beyond the blacklists. Our contributions are as follows:

- We propose a fog-based firewall system for information-centric IoV, which utilises content awareness and edge-distributed computation of fog to construct information-centric IoV edge isolation, and thus to prevent threatening traffic and requests entering information-centric IoV.
- The semantic reasoning scheme to mine potential threats in packets is proposed with knowledge graph. Firstly, a context selection scheme based on semantics is proposed to guide the reasoning direction of interest packets. Moreover, weighted semantic reasoning is designed to reason the threatening relations and knowledge with interest packets and thus to supplement blacklists with potential threats in both interest and data packets. Based on the knowledge constructed from reasoning, the proposed semantic reasoning scheme perceives penetrated and obfuscating content threats, and configures customised defence policies with distinct interests.
- The proposed information-centric IoV firewall system can defend against both content threats and network layer threats in information-centric IoV. The considered threats cover network layer attacks of anomaly requests, packet tempering and falsifying, content as well as cache attacks of information leakages, illegal and obtucating data.

The remaining of this paper is organised as follows: Section 2 describes the research related to this paper. The design principles of content threat-aware fog-firewall system are introduced in Section 3. In Section 4, the firewall architectures and workflow are designed. Section 5 presents the context based interest semantic reasoning for threat-aware policy customisation. Simulations and the preliminary results are provided in Section 6. Finally, we draw our conclusion in Section 7.

2 Related work

In recent years, there have been frameworks based on information-centric IoV to combine ICN and IoV. Researchers have applied ICN into IoV to overcome the above disadvantages of the IP-based end-to-end communication. Moreover, it is compatible with the fact that IoV network services are typically dominated by content distribution.

In [2], Li et al. proposed a novel crowd sourced VCCN framework to provide secure and efficient information distribution for IoV. It supports vehicles to crowdsource their cache resources and radio links then to distribute collaborative content. Besides, the authors in [9] designed a one-hop large-size content distribution mechanism for IoV, which applies the emerging NDN to V2V communications. It enables services including nearby requests, intelligent responses, multi-source supplies and breakpoint resume to be fulfilled. In [10], a hierarchical location based content naming mechanism was proposed for IoV. Besides, the authors also designed a distributed mobility management mechanism via taking directions of movement in the content names into consideration. However, the aggregation of the proposed location based mechanism will be declined due to the mobility of vehicles.

Recently, some researchers have focused on the security and communication study of IoT [11–13] including IoV [14], Smart Grid [15–18] and SDN [19] as well. However, studies on information-centric IoV firewalls are still in the infancy. In [20], a study of CCN security requirements is introduced and the first CCN firewall with syntax is designed. However, filtering is conducted for isolated packets without consideration of context, leaving vulnerabilities of latent and continuous attacks. Moreover, the proposed semantic analysis focuses on name syntax while neglecting relations between packet and blacklist. In [21], the distributed context-aware firewall takes account of the specialised knowledge of protected services to decrease processing delay. However, it fails to be applied in information-centric IoV scenarios for distinct network protocols and architecture. In [22], a simple filter technique based on web URL statistics for firewalls is
proposed to detect anomalous information-centric IoV names. The authors in [23] present a novel system for detecting and avoiding poisoned content, which leverages the verification work that users must do anyway. The above studies propose efficient protection policies of information-centric IoV, while little attention is paid to constructing the firewall system for it.

For the employing of information-awareness [24], ICN will be widely used in 5G, Smart Grid and Vehicle to Grid [25–27]. An efficient access control framework for information-centric network (ICN) is proposed in [28], which allows legitimate users to access and utilises the cached content directly without verification/ authentication. The work in [29] designs a secure content distribution architecture for CCN that is based on proxy re-encryption. Misra et al. create a secure content delivery framework especially applicable for mobile devices in [30]. Zhang et al. [31] proposed a cooperative random interest propagation mechanism to prevent content-client link ability. Existing studies provide good support to resolve various information-centric IoV content threats in network layer. However, information-centric IoV content threats penetrated in semantics and ambiguity are neglected.

Fog computing has been widely applied in next generation networks [32–35]. Studies involve optimising computation resources of the application layer [36], making use of content awareness to reassign services [37]. Considering the advantages of content awareness and edge distribution, fog computing can provide great benefits to information-centric IoV firewalls in the performance of threat-aware filtering and semantic reasoning to construct edge defence isolation.

To sum up, much more attention needs to be paid to information-centric IoV firewalls with fog computing on the edge of networks. Interest semantic analysis to mine implicit threats in content is of vital importance for content obfuscation and penetration. Customised policy configuration with user and threat awareness should be considered for distinct traffics.

3 Design principles of content threat-aware fog-firewall system

This section showed the overview of proposed filtering policies and prevented information-centric IoV threats. Two types of the considered network boundaries were designed where proposed fog-firewalls were placed.

3.1 Defence principles and categories of threat defence policies

The threats we considered for information-centric IoV security isolation are divided into two aspects: content threats and network layer threats. Content threats refer to information leakage, illegal or unexpected content transmission, and content obfuscation. Content threats also cover cache attacks that attackers fulfil the cache nodes with invalid or unexpected content. For the network layer, threats include anomaly requests, spoofing that attackers monitor, block, temper and falsify the packets [6, 7] and illegal participants. Table 1 shows the considered security threats and our proposed filtering detection objects. The proposed information-centric IoV firewall system aims to provide a security isolation system defending both content and network layer threats. The filtering principles are designed as follows.

- **Data integrity checking**: Detect if data is tempered or producers in blacklists send responses so that illegal or uncorrelated content is sent out with requested names.

- **Content filtering**: On the one hand, with the analysis of interest names, anomaly requests of illegal or malicious invalid data are prevented from entering information-centric IoV. For data packets, illegal and unexpected content of data packets are blocked with the blacklists on the boundaries, which avoids invalid data being cached. On the other hand, security content isolation between different networks is under consideration, which includes unauthorised access and information leakage.

**Customised filtering policy**: Besides the inherent blacklist, semantic reasoning of information-centric IoV interest is conducted to mine potential threatening knowledge, and then supplement blacklists for customised policy configuration. Moreover, components of information-centric IoV packets, including name components, selectors, ForwardingHint, producer signatures and content [5] are extracted as filtering policies to exclude unexpected sources and content of response packets. Thus the proposed firewalls support consumers to dynamically configure filter policies.

3.2 Content threat awareness with interest semantic reasoning

Considering the inherent ambiguity in human-readable language, content obfuscation and penetration are novel information-centric IoV threats. Malicious nodes send implicitly forbidden content as data packets that are obfuscated and correlated with requested names in some semantic dimensions. It can lead to the failure of recognition with the inherent firewall blacklists. As for interest packets, the names of the requested content may have potential threatening relations with entities in blacklists while do not appear explicitly.

Therefore, we pay special attention to content obfuscation attacks for its unique need of semantic analysis. For content threats with semantic ambiguity and penetrated illegality, an interest semantic reasoning is designed with knowledge graph to mine potential threatening knowledge through names of interest packets. Threats to be mined include relations between entities in requested content attributes and in blacklists, and threatening entities composting the potential relations between packets and blacklists. Based on semantic reasoning of each information-centric IoV interest packets, we detect the implicit illegal features of interest and then predict obfuscating threatening responses to be excluded.

**Furthermore**, for latency and continuously attacks, context communication related with the pending packets in content, names or producers are taken into consideration, which guides the reasoning direction to focus on the semantic dimensions of requests. The above mentioned schemes are introduced in Section 5.

3.3 Edge defence isolation of proposed fog-firewall system

As shown in Fig. 2, IoV nodes are deployed at the edge and bottom layer of the whole networks as terminals, including vehicles, distinct kinds of sensors on vehicles, passengers and other infrastructures. They act and serve as the data creator to generate data. Terminals do not need to be equipped with firewalls because many of them cannot bare the required computation and storage for a firewall service. Besides, by deploying firewalls in the upper fog nodes, it helps the firewall systems conduct more comprehensive analysis for traffic and perceive threats with not only a single node but the covered local networks.

Fog computing nodes are deployed at the middle layer in the information-centric IoV. They act as a bridge and also a security isolation system. In this paper, two kinds of edge defence isolation are introduced based on different attacks for vehicles and Intranet. On the one hand, terminal isolation is constructed between information-centric IoV terminals and networks. Taking advantages of the conventional terminal firewalls, it can isolate illegal communication between terminals and information-centric...
IoV. Terminal isolation is deployed logically between terminals and next hop of information-centric IoV nodes to prevent malicious data from entering information-centric IoV. As for the physical deployment, it is placed on fog nodes that achieve seamless coverage of terminal geographically. Each terminal is in a jurisdiction of one corresponding fog-firewall. Moreover, due to the computation of fog nodes, the burden for terminals of configuring firewalls is sharply reduced.

On the other hand, network isolation that prevents malicious packets from spreading between different information-centric IoV networks is established. Traditional network boundaries that connect different levels of security networks are considered here to isolate communication between Intranet and Extranet. Thus, information-centric IoV network isolation is deployed on fog nodes beside gateways. A physical gateway will route both inside and outside arriving packets to the network firewall after process.

ICN nodes are the components of the top layer. They act as the core net and provide addressing, routing and caching functions. Fog nodes are responsible for building the isolation system between the other two layers.

We describe the attack and defence scenarios in Fig. 2 as well. The scenarios are classified according to attacking sources. For malicious requests, we consider both anomaly requests from edge terminals and unauthorised requests between LANs. The formal can be detected by signatures and packet names in terminal firewalls, thus they cannot access the network. This is because signature and name component include the identity and content abstract information which cover the threat information. The latter refers to unpermitted access towards external LANs and it cannot be detected by terminal firewalls. However, on the gateways, firewalls can detect and block it according to packet names. Therefore, unauthorised access can be prevented outside the network. On the other hand, for malicious responses, we give the defence scenarios of threats considered in Section 3.1. Data attacks including cache, obfuscation, falsifying and leakage could be blocked by parsing the signature, name and data component of packets to check the security of data. Moreover, illegal data producers can be detected by the signature.

4 Proposed content threat-aware fog-firewall system with customised policies

The proposed content threat-aware fog-firewall system for information-centric IoV consists of terminal isolation fog-firewalls (TI2Fs) and network isolation fog-firewalls (NI2Fs). TI2Fs block illegal requests as well as data of terminals entering information-centric IoV. NI2Fs detect Extranet illegal access and prevent Intranet data leakage. In this section, we design both the proposed TI2F, NI2F and the workflow of the firewall defence operations.

4.1 Content threat-aware TI2Fs

The TI2F is designed based on the hierarchical fog model proposed in [34] as shown in Fig. 3. For packets sent from terminals, they are firstly routed to a TI2F in monitoring layer, and parsed into preprocessing layer to extract components. Interest packets are then analysed in the policy layer to generate customised filter policies for both response and itself. The filter layer conducts block with inherent and customised filter policies. The security layer attaches customised policies to interest, thus policies could be reused when both response and itself. The filter layer conducts block with inherent and customised filter policies. The security layer attaches customised policies to interest, thus policies could be reused when conferring firewalls in the monitoring layer. The routing layer then forwards the new packet to the next hop.

The monitoring layer is to retrieve and then record all the communication behaviours of each information-centric IoV nodes. Monitored objects include terminals, network topology, activities and resources. Therefore, TI2F has access to all the requests, responses and device attributes of each information-centric IoV node, which helps perceive user behaviour and context traffic. Data obtained in the monitoring layer is stored in two databases. Device attributes and location information are placed in the network database. The communication activities are recorded in the log database.

The preprocessing layer parses packets and extracts the components of each packet. Components extracted from interest packets include packet names, selectors and ForwardingHint [37]. The selectors refer to specified publisher keys and excluding information for names of responses. And ForwardingHint represents an authorisation list of forwarding objects. For data packets, components of name, signature, sighed information and content are extracted into the database.

The policy layer configures customised filtering policies for distinct packets dynamically. It consists of three modules: correlated context analysis, semantic reasoning and user-configured policy extraction. For interest packets, correlated context is selected and relevance weight matrix is calculated to guide the reasoning direction. Then the policy layer conducts

---

**Fig. 2** Attack and defence scenarios of fog-based information-centric IoV firewalls

**Fig. 3** Structure of proposed TI2Fs

*This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)*
Intranet and Extranet unauthorised access, as shown in Fig. 4. The difference between content filter of interest and data packets is the falsity content attacking.

In the filtering layer, for interest packets, requests are firstly acquired by TI2Fs before entering information-centric IoV. TI2Fs can conduct filtering on data packets and decide whether to block or allow it to enter information-centric IoV. They are parsed similarly with that in terminal firewalls. The packet database in a TI2F where all the communication records are parsed into key entities upon name, selector and the context correlated with the pending packet is collected according to the filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

The monitoring layer acquires packets from both Intranet and Extranet. They are parsed similarly with that in terminal firewalls. The policy layer contains the communication record analysis and semantic reasoning module. When receiving an interest packet, related communication content is calculated and fetched from the packet database. Then semantic reasoning of the blacklist, interest names and previous communication records are conducted to find possible potential illegal entities. The filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

### 4.3 Defence workflow of fog-firewall system

The workflow of our proposed firewall system is shown in Fig. 5. For interest packets, requests are firstly acquired by TI2Fs before entering information-centric IoV. TI2Fs can conduct filtering on data packets and decide whether to block or allow it to enter information-centric IoV. Similarly, with the case of interest packets, no more TI2Fs will inspect the data packet on the forwarding path for data in the information-centric IoV has inspected on the entry. In the case of internetwork communication, when an interest packet is routed to another private net, the NI2F conducts access authority control to prevent illegal access. For data packets, the Intranet physical gateway sends it to the NI2F inspecting information leakage. When routed to the requestor's private network, access authority control is conducted again to avoid malicious node sending illegal or unexpected content.

### 5 Interest semantic reasoning for threat-aware policy customisation of proposed firewalls

We first design a context selection algorithm: Communication context correlated with the pending packet is collected according to content similarity and the relevance weights are calculated. And two reasoning algorithms of knowledge graph are modified and applied to reason the potential relations and entities between requests and entities in blacklists.

#### 5.1 Threat-aware direction guidance with context

The packet database in a TI2F where all the communication packets are stored can hold N records for each covered terminal and N is adjusted with calculation ability of each fog node. The database adopts first-in-first-out principle. All the packets in the database are parsed into key entities upon name, selector and the context correlated with the pending packet is collected according to the filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

The monitoring layer acquires packets from both Intranet and Extranet. They are parsed similarly with that in terminal firewalls. The policy layer contains the communication record analysis and semantic reasoning module. When receiving an interest packet, related communication content is calculated and fetched from the packet database. Then semantic reasoning of the blacklist, interest names and previous communication records are conducted to find possible potential illegal entities. The filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

The monitoring layer acquires packets from both Intranet and Extranet. They are parsed similarly with that in terminal firewalls. The policy layer contains the communication record analysis and semantic reasoning module. When receiving an interest packet, related communication content is calculated and fetched from the packet database. Then semantic reasoning of the blacklist, interest names and previous communication records are conducted to find possible potential illegal entities. The filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

The semantic reasoning between request names and blacklists. Thus the reasoning constructs security knowledge policy and finds the related new entities to expand the blacklist. Moreover, user-configured filtering policies are generated by combining extracted packet components. Selector components are used to restrict producer identities and to exclude packet names while forwardingHint is used to control packet transmitting. All polices for one specific interest packet are called interest-specified policies (ISPs).

In the filtering layer, for interest packets, integrity and validity are checked to prevent requests and ISPs being tampered by malicious nodes. Besides, content matching based on the blacklist is conducted to prevent anomaly requests. While for data packets received, integrity and validity authority, user-configured policy matching and content filter are implemented orderly. The difference between content filter of interest and data packets is the processed objects. Only the content names in interests are filtered for malicious request detection whereas the whole content in data packets are filtered to prevent illegal, unexpected, obfuscated and falsity content attacking.

The security layer consists of two modules: a tagging module and an encryption module. The tagging module adds a new ISP tag in interest packets from the policy layer and expands the blacklist with reasoned threatened entities. Thus, other information-centric IoV nodes can acquire ISPs when caching the interest. Related TI2Fs can obtain ISPs in the monitoring layer as well and place them in the blacklist database. The encryption module conducts encryption for a tagged interest. The routing layer then forwards new packets to the next hop into information-centric IoV.

#### 4.2 Content threat-aware NI2Fs

A NI2F is designed to prevent data leakage from a protected Intranet and Extranet unauthorized access, as shown in Fig. 4. Unlike the TI2F, there is no security layer in network firewalls. Interest packets routed to the network boundary have already been parsed in their TI2Fs to generate ISPs.

The workflow of our proposed firewall system is shown in Fig. 5. For interest packets, requests are firstly acquired by TI2Fs before entering information-centric IoV. TI2Fs can conduct filtering on data packets and decide whether to block or allow it to enter information-centric IoV. Similarly, with the case of interest packets, no more TI2Fs will inspect the data packet on the forwarding path for data in the information-centric IoV has inspected on the entry. In the case of internetwork communication, when an interest packet is routed to another private net, the NI2F conducts access authority control to prevent illegal access. For data packets, the Intranet physical gateway sends it to the NI2F inspecting information leakage. When routed to the requestor's private network, access authority control is conducted again to avoid malicious node sending illegal or unexpected content.

5 Interest semantic reasoning for threat-aware policy customisation of proposed firewalls

We first design a context selection algorithm: Communication context correlated with the pending packet is collected according to content similarity and the relevance weights are calculated. And two reasoning algorithms of knowledge graph are modified and applied to reason the potential relations and entities between requests and entities in blacklists.

#### 5.1 Threat-aware direction guidance with context

The packet database in a TI2F where all the communication packets are stored can hold N records for each covered terminal and N is adjusted with calculation ability of each fog node. The database adopts first-in-first-out principle. All the packets in the database are parsed into key entities upon name, selector and the context correlated with the pending packet is collected according to the filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

The monitoring layer acquires packets from both Intranet and Extranet. They are parsed similarly with that in terminal firewalls. The policy layer contains the communication record analysis and semantic reasoning module. When receiving an interest packet, related communication content is calculated and fetched from the packet database. Then semantic reasoning of the blacklist, interest names and previous communication records are conducted to find possible potential illegal entities. The filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

The monitoring layer acquires packets from both Intranet and Extranet. They are parsed similarly with that in terminal firewalls. The policy layer contains the communication record analysis and semantic reasoning module. When receiving an interest packet, related communication content is calculated and fetched from the packet database. Then semantic reasoning of the blacklist, interest names and previous communication records are conducted to find possible potential illegal entities. The filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

The monitoring layer acquires packets from both Intranet and Extranet. They are parsed similarly with that in terminal firewalls. The policy layer contains the communication record analysis and semantic reasoning module. When receiving an interest packet, related communication content is calculated and fetched from the packet database. Then semantic reasoning of the blacklist, interest names and previous communication records are conducted to find possible potential illegal entities. The filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

The monitoring layer acquires packets from both Intranet and Extranet. They are parsed similarly with that in terminal firewalls. The policy layer contains the communication record analysis and semantic reasoning module. When receiving an interest packet, related communication content is calculated and fetched from the packet database. Then semantic reasoning of the blacklist, interest names and previous communication records are conducted to find possible potential illegal entities. The filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.

The monitoring layer acquires packets from both Intranet and Extranet. They are parsed similarly with that in terminal firewalls. The policy layer contains the communication record analysis and semantic reasoning module. When receiving an interest packet, related communication content is calculated and fetched from the packet database. Then semantic reasoning of the blacklist, interest names and previous communication records are conducted to find possible potential illegal entities. The filtering layer blocks Extranet requests beyond the access authority. Besides, data packets sent to Extranet are detected with blacklists for information leakage. The routing layer then forwards packets into information-centric IoV.
\[ J_{ij} = \frac{S^h_i \cap S^e_j}{S^h_i \cup S^e_j} = \frac{\sum_{k=1}^M s^h_k \cdot s^e_j}{\sum_{k=1}^M (s^h_k + s^e_j) - \sum_{k=1}^M s^h_k \cdot s^e_j} \]  

(2)

Then the relevance value of \( p^i \) and \( p^j \) is normalised as

\[ r^{i,j} = \frac{J_{ij}}{\sum_{j=1}^N J_{ij}} \]  

(3)

We select top \( K \) packets with highest relevance weight as the correlated context. The relevance matrix \( R_{K \times K} = \{ r^{i,j} \} \) is established with the relevance weight \( r^{i,j} \) of selected packets in descending order. Entities of the selected packets form another entity matrix \( S_{K \times K} = \{ s^j \} \) where the column vectors \( S^j \) are sorted in descending order of \( r^{i,j} \).

\[ K = \left[ N \sum_{j=2}^N (J_{ij} - J_{i1}) \right] \]  

(4)

The number of \( K \) varies with the average degree of context relevance and the calculation ability of fog. A higher overall relevance of context with the pending packet leads to more correlated packets and their key entities to be selected for reasoning. And a stronger fog node can deal with more correlated packets as well. All the key entities of the selected context form a new collection \( E_c \). The relevance weight \( r^{i,j} \) of a selected entity to be reasoned with blacklists is

\[ r^{e_l} = \sum_{i=1}^K s^h_i \cdot R_{i \times K}^T \]  

(5)

where \( s^h_i \) is the \( i \)th row vector of \( S^h_{K \times K} \). Therefore, we select top \( K \) correlated packets and the relevance weight \( r^{e_l} \) of their key entities, which are related with the request content.

5.2 Interest semantic reasoning for security knowledge policy

Semantic reasoning included two aspects: To reason the potential threatening relation between the request and entities in blacklists and to mine the sensitive entities that can composite a multiple-step relation between the requests and blacklists.

To mine possible threatened relations between requests and blacklists, we introduced and modified the TransA model, which utilises the metric learning to learn the relation between entities in knowledge graph [38]. TransA proposes a loss metric with a score function (6) to estimate relations of entity pairs and a weight matrix to weight specific feature dimensions during reasoning training, where \( e_0, r, e_i \) refer to the head entity, relation and tail entity in the knowledge graph

\[ F_h(e_0, e_i) = \left\langle p_0 + r - e_i \right\rangle^T W_i \left( p_0 + r - e_i \right) \]  

(6)

We modified the weight matrix \( W_i \) with the semantics of correlated context in (7). \( \Delta \) is a dataset comprised of existing triples \( (e_0, r, e_i) \) in knowledge base (KB), \( \Delta \) denotes the negative samples constructed with false triples. When the selected training triple contains the entity in context collection \( E_c \), the relevance value is applied in \( W_i \) to reduce the score (6) in loss metric with a factor of \( e^{-\gamma} \) during training. Then semantic dimensions of the selected context differ from general dimensions in \( W_i \) which results in a faster reduction of loss metric

\[ W_i = \sum_{(e_0, r, e_i) \in \Delta} A \left\langle \left[ e_0^i + r^j - e^j \right] \left[ \left[ e_0^i + r^j - e^j \right] \right] \right\rangle - \sum_{(e_0, r, e_i) \in \Delta} A \left\langle \left[ e_0^i + r - e^j \right] \left[ \left[ e_0^i + r - e^j \right] \right] \right\rangle \]  

(7)

where

\[ A = \begin{cases} e^{-\gamma}, & \text{if } e_0^i \text{ or } e_i^j \in W_e \\ 1, & \text{others} \end{cases} \]  

(8)

We took KB triples as learning samples to optimise parameters in loss metric. The loss function value continuously reduced through learning, and the entity vector and relation matrix can better reflect the semantic information of entities and relations. Relations with the similar semantics with the context and requests were learnt as a result.

We modified the PTransE model to find the potential sensitive entities to composite a relation between requests and blacklists, which are likely to be the obfuscating objects in the responded data packets. PTransE takes multiple-step relation paths into consideration for representation learning [39], the score function of multiple steps in loss metric is defined in (9). We trained the model with the same datasets of TransA to optimise the loss metric thus obtaining multiple-step relations of entity pairs \( P^m(e_0, e_i) = \{ p_0^1, \ldots, p_0^N \} \), where each relation path is \( p^m = (r_1, \ldots, r_t) \). Middle entities represented as \( E_0 \) can be mined as a result.

\[ F_h(e_0, P^m, e_i) = \frac{\sum_{e_0 \in E_0} R(p^m | e_0, e_i)}{\sum_{e_0 \in E_0} R(p^m | e_0, e_i)} \]  

(9)

To guide the relation path direction, we modified the resource function \( R(p^m | e_0, e_i) \) which measures the resource flowed from \( e_0 \) to \( e_i \) as the path reliability in (10). When the middle entity \( e_0 \) \( e_i \in W_e \), the resource allocated to the next entity is weakened by the relevance weight resulting in a lower loss function in learning. And \( A \) is defined in (8)

\[ R(p^m | e_0, e_i) = \sum_{e_0 \in E_0} A \]  

(10)

Therefore, for each request, the blacklists are expanded with the reasoned middle entities. A response will be blocked when the middle entities are detected in the packet content.

6 Simulation and analysis

In this section, we simulate the proposed fog computing based information-centric IoV firewall system and semantic reasoning based threat awareness with ndSIM. ndSIM is a NS-3 based MN simulator (version ns3SIM 2.2). NS-3 (Network Simulator V3) is a simulating software of discrete events for the NDN research which has granted a license of GNU GPLv2 [40]. Simulations on cache hits, defence accuracy and end to end communication delay are conducted. And the results show that our proposed fog firewall performs well in terms of scalability, defence accuracy and efficiency.

6.1 Scalability evaluation of threat-aware fog-firewall

In the information-centric IoV, there are always new vehicles and pedestrians accessing and leaving the vehicle network casually with the change of time-domain. Thus, to verify the scalable capacity of the proposed fog computing based information-centric IoV firewall system as well as the semantic reasoning based threat awareness schema, the cache hit of ICN nodes is evaluated with different network scales. The reason why we choose cache hits as the evaluation demission is because in information-centric IoV, ICN nodes will store the forwarding content into their cache.
information-centric IoV and a fog firewall-free information-centric IoV will be cached in routers. And the results prove that our proposed fog firewall-equipped information-centric IoV are lower and grow slower than that of the firewall-free information-centric IoV. This is because anomaly requests, illegal and unexpected data are blocked on the firewall so that they cannot enter information-centric IoV or to be cached in routers. And the results prove that our proposed fog firewall model as well as semantic reasoning algorithm can filter sensitive threat content outside the network. Besides, our model performs well as the network scale varies.

6.2 Defence accuracy evaluation of threat-aware fog-firewall

It is important for the firewall to perceive, mine and then defend both explicit and inherent threat information. To evaluate the defence-accuracy of our model, we define a variable which is unknown threat rates (UTRs).

\[
UTR = \frac{|E_h|}{|E_h| + |E_i| + |E_s|}
\]

And \(E_h, E_i\) present the entities of Knowledge Graph. Here we choose the WN18 which is a dataset WordNet as the Knowledge Graph. We randomly select 50 triples in WN18, including a head entity \(E_h\), a tail entity \(E_t\). And then all the relation attributes which connect the two entities, are mined using our semantic reasoning algorithm. The tail entity collection \(E_t\) is listed in blacklists as known threats and the heads \(E_h\) act as implicit threats which should be mined and blocked by reasoning. In addition, we utilise Word2Vec to train synonyms or similar entities \(E_t\) of the head \(E_h\) from Wikicorpus and they act as correlated context to guide the reasoning from \(E_h\) to \(E_t\).

Defence accuracy of the proposed firewall with semantic reasoning is simulated. We evaluated defence accuracies with different UTRs which are 30, 50, and 70%. Nodes in the simulated information-centric IoV randomly request content in the three mentioned collections. Simulations of each network topology were conducted 100 times and then we calculate the average accuracy. As shown in Fig. 7, firewallss with semantic reasoning always perform higher defence accuracy than those whose blacklists are fixed with \(E_t\) only as information-centric IoV expands. The accuracy rises because with more correlated context sent to firewalls, more \(E_h\) as well as \(E_t\) are mined to be relevant with \(E_t\) and thus are blocked, which shows the proposed semantic reasoning performed well in perceiving threats.

6.3 Efficiency evaluation of threat-aware fog-firewall

When the firewall conducts analysis and blocking, the network transmission delay will get influenced. This is because a fog firewall needs to cover more than one terminal, and when the traffic becomes extremely enormous, content needs to wait in line for getting filtered. Therefore, we evaluate the processing efficiency of our defence isolation provided by the fog-firewall model.

We choose end-to-end communication delay to evaluate the influence that the firewall produced on the information-centric IoV communication transmission. The end to end communication delay is calculated from the requester send the package to the responder receive it, which includes the transmission delay between each routing nodes and the processing delay both firewalls and routing nodes. We simulate two scenarios that firewalls are deployed on fog nodes and on terminals directly. For both scenarios, two networks are designed with the same topology, requests and nodes. Therefore, the influence of processing delay on routers can be neglected because the routing time is the same for both scenarios. By changing the frequency of terminals requesting content, we can get different delays. The average communication delay has been counted, and a comparison is shown in Fig. 8. As the interest request rate rises, fog-based firewalls present higher processing efficiency than firewalls on terminals. The average delay in fog-firewalls increases more slowly and is almost lower than that of terminal-deployed firewalls. This is because although there might be queuing phenomenon in fog firewall, yet fog nodes provide crucial computation and awareness resources to help firewalls perceive and predict user behaviour, reason defence policies and conduct blocking while general terminals cannot support light weight but accurate defence especially when the network traffic becomes huge. In addition, firewalls deployed on terminals cannot perceive some threats that have been found by nearby nodes, which waste repeated time to analyse.
threatened knowledge from interest names and contents. The entering information-centric IoV and unexpected access between different private networks. An interesting semantic reasoning approach with context was proposed to mine potential relations and threatened knowledge from interest names and contents. The simulation results showed that the proposed fog based information-centric IoV firewall system could provide scalable, accurate and efficient isolation defence. This work is significant to improve the security of information-centric IoV.

8 Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant nos. 61431008 and 61571300).

9 References

[1] Amadeo, M., Campolo, C., Molinaro, A.: ‘Information-centric networking for connected vehicles: a survey and future perspectives’, IEEE Commun. Mag., 2016, 54, (2), pp. 98–104
[2] Li, C., Gong, S., Wang, X., et al.: ‘Secure and efficient content distribution in crowdsourced vehicular content-centric networking’, IEEE Access, 2018, 6, pp. 5727–5739
[3] Wang, M., Wu, J., Li, G., et al.: ‘Toward mobility support for information-centric IoV in smart city using fog computing’, IEEE Int. Conf. Smart Energy Grid Engineering (SEGEN), Oshawa, Canada, August 2017, pp. 357–361
[4] Li, Q., Sandhu, R., Zhang, X., et al.: ‘Mandatory content access control for privacy protection in information-centric networks’, IEEE Trans. Dependable Secure Comput., 2017, 14, (5), pp. 494–506
[5] Jacobson, V., Smetters, D., Thornton, J., et al.: ‘Networking named content’, Int. Conf. Emerging Networking Experiments and Technologies, Rome, Italy, December 2009, pp. 1–12
[6] Tourani, R., Misra, S., Mick, T., et al.: ‘Security, privacy, and access control in information-centric networking: a survey’, IEEE Commun. Surv., 2020, 20, (1), pp. 566–600
[7] AbdAllah, E., Hassanien, H., Zulkernine, M., et al.: IEEE Commun. Surv., 2015, 17, (3), pp. 1441–1454
[8] Mukherjee, M., Sha, L., Wang, D., et al.: ‘Survey of fog computing: fundamental, network applications, and research challenges’, IEEE Commun. Surv., 2018, 20, (3), pp. 1–30
[9] Deng, G., Shi, L., Li, R., et al.: ‘Efficient inter-vehicle internet content distribution based on named data’, Vehicular Technology Conf., Boston, MA, September 2015, pp. 1–5
[10] Yan, Z., Zeadally, S., Park, Y.: ‘A novel vehicular information network architecture based on named data networking’, IEEE Internet Things J., 2014, 1, pp. 525–532
[11] Wu, J., Luo, S., Wang, S., et al.: ‘NLES: a novel lifetime extension scheme for safety-critical cyber-physical systems using SDN and NFV’, IEEE Internet Things J., 2018, pp. 1–13, doi: 10.1109/JOT.2018.2870294
[12] Wu, C., Liu, Z., Zhang, D., et al.: ‘Spatial intelligence towards trustworthy vehicular IoV’, IEEE Commun. Mag., 2018, 56, (10), pp. 22–27
[13] Zhang, Y., Yu, R., Yao, W., et al.: ‘Home M2M networks: architectures, standards, and QoS improvement’, IEEE Commun. Mag., 2011, 49, (4), pp. 42–48
[14] Zhu, Z., Wang, Z., Chu, Z., et al.: ‘Robust energy harvest balancing optimization with V2X-SWIPT over MISO secrecy channel’, Comput. Netw., 2018, 137, pp. 61–68
[15] Wu, J., Ota, K., Dong, M., et al.: ‘Big data analysis-based security situational awareness for smart grid’, IEEE Trans. Big Data, 2018, 4, (3), pp. 408–417
[16] Guan, Z., Li, J., Wu, L., et al.: ‘Achieving efficient and secure data acquisition for cloud-supported internet of things in smart grid’, IEEE Internet Things J., 2017, 4, (6), pp. 1934–1944
[17] Maharjan, S., Zhu, Q., Zhang, Y., et al.: ‘Dependable demand response management in the smart grid: a Stackelberg game approach’, IEEE Trans. Smart Grid, 2013, 4, (1), pp. 120–132
[18] Zhang, Y., Yu, R., Nekovee, M., et al.: ‘Cognitive machine-to-machine communications: visions and potentials for the smart grid’, IEEE Netw., Mag., 2012, 26, (3), pp. 6–13
[19] Wu, J., Dong, M., Ota, K., et al.: ‘Big data analysis-based secure cluster management for optimized control plane in software-defined networks’, IEEE Trans. Netw. Serv. Manage., 2018, 15, (1), pp. 27–38
[20] Goergen, D., Cholez, T., Francois, J., et al.: ‘A semantic firewall for content-centric networking’, IFIP/IEEE Int. Symp. Integrated Network Management (IM 2013), Ghent, Belgium, May 2013, pp. 478–484
[21] Varas, C., Hirsch, T.: ‘Self-protection through collaboration using D-CAF: a distributed context-aware firewall’, Third Int. Conf. Emerging Security Information, Systems and Technologies, Athens/Glyfada, Greece, June 2009, pp. 179–184
[22] Kondo, D., Silverston, T., Tode, H., et al.: ‘Name anomaly detection for ICN’. IEEE Int. Symp. Local and Metropolitan Area Networks (LANMAN), Rome, Italy, June 2016, pp. 1–4
[23] Dillenbetta, S., Papadopoulos, C.: ‘Mitigating poisoned content with forwarding strategy’. IEEE Conf. Computer Communications Workshops (INFOCOM WKSHPS), San Francisco, CA, April 2016, pp. 164–169
[24] Xylomenos, G., Ververidis, C., Siris, V., et al.: ‘A survey of information-networking research’, IEEE Commun. Surv., 2014, 16, (2), pp. 1024–1049
[25] Ravindran, R., Chakraborti, A., Amin, S.O., et al.: ‘5G-ICN: delivering ICN services over 5G using network slicing’, IEEE Commun. Mag., 2017, 55, (5), pp. 101–107
[26] Wang, K., Li, J., Wu, J., et al.: ‘QoS-predicted energy efficient routing for information-centric smart grid: a network calculus approach’, IEEE Access, 2018, 6, pp. 52867–52876
[27] Li, G., Wu, J., Li, J., et al.: ‘Battery status sensing software-defined multicast for V2G regulation in smart grid’, IEEE Sens. J., 2017, 17, (23), pp. 7838–7848
[28] Misra, S., Tourani, R., Natividad, F., et al.: ‘Accomf: an access control framework for leveraging in-network caching in the ICN-enabled wireless edge’, IEEE Trans. Dependable Secure Comput., 2017, 16, (1), pp. 5–17
[29] Wood, C., Uzun, E.: ‘Flexible end-to-end content security in CCN’. IEEE 11th Consumer Communications and Networking Conf. (CCNC), Las Vegas, NV, January 2014, pp. 858–865
[30] Misra, S., Tourani, R., Majid, N.: ‘Secure content delivery in information-centric networks: design, implementation, and analyses’. Proc. ACM SIGCOMM Workshop on Information-centric Networking, Hong Kong, China, August 2013, pp. 73–78
[31] Zhang, P., Li, Q., Li, P.: ‘Achieving content-oriented anonymity with CRISP’, IEEE Trans. Dependable Secure Comput., 2017, 14, (6), pp. 578–590
[32] Wu, J., Dong, M., Ota, K., et al.: ‘FCSS: fog-computing-based content-aware filtering for security services in information-centric social networks’, IEEE Trans. Emerg. Top. Comput., 2017, pp. 1–12, doi: 10.1109/TETC.2017.2747158
[33] Li, G., Wu, J., Li, J., et al.: ‘Service popularity-based smart resources partitioning for fog computing-enabled industrial internet of things’, IEEE Trans. Ind. Inf., 2018, 14, (10), pp. 4702–4711, doi: 10.1109/TII.2018.2845844
[34] Zeng, D., Gu, L., Guo, S., et al.: ‘Joint optimization of task scheduling and image placement in fog computing supported software-defined embedded system’, IEEE Trans. Comput., 2016, 65, (12), pp. 3702–3712
[35] Guan, Z., Zhang, Y., Wu, L., et al.: ‘APPA: an anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT’, J. Netw. Comput. Appl., 2019, 129, pp. 82–92
[36] Dong, M., Ota, K., Liu, A.: ‘Preserving source-location privacy through redundant fog loop for wireless sensor networks’. IEEE Int. Conf. Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, England, October 2015, pp. 1835–1842
[37] Aazam, M., Huh, E.: ‘Fog computing: the cloud-IoT-to-IoT middleware paradigm’, IEEE Potentials, 2016, 35, (3), pp. 40–44
[38] Xiao, H., Huang, M., Hao, Y., et al.: ‘Transa: an adaptive approach for knowledge graph embedding’, 2015, arXiv preprint arXiv:1509.05490
[39] Lin, Y., Liu, Z., Luan, H., et al.: ‘Modeling relation paths for representation learning of knowledge bases’, 2015, arXiv preprint arXiv:1506.00379
[40] Han, W., Dong, M., Ota, K., et al.: ‘SD-OPTS: software-defined on-path time synchronization for information-centric smart grid’. 2017 IEEE Global Communications Conf., Singapore, December 2017, pp. 1–6