STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LXI**, Number 2, 2016

COALGEBRAIC SPECIFICATION OF NETWORK INTRUSION SIGNATURES

JÁN PERHÁČ, AND DANIEL MIHÁLYI

Abstract. This paper deals with a formal description of the intrusion detection system behavior using coalgebras and coalgebraic specification of network intrusion signatures. We show how intrusion detection system's real network intrusion signatures can be specified by coalgebraic signatures and we propose our approach related to modeling program system behavior as coalgebra for polynomial endofunctor over category of the infinite stream of packets.

1. Introduction

Category theory [1] appears to be a suitable mathematical structure designated for a formal description of infinite data structures, i.e. streams. In the computer networks area, we define the packets passing in a particular network segment as infinite stream. Our contemporary effort is related to explore how network intrusion signatures can be described formally using coalgebraic signatures. Here we differentiate two types of signatures: the first one is packet signature that expresses packet attributes and packet behavior from the intrusion detection system (IDS) point of view. The second one is Snort's (a real implementation of lightweight network IDS) intrusion signatures related to particular network anomalies that consist from anomaly structure and its symptoms.

So far, there are several works using formal methods to achieve state dynamic IDS behavior description based on the Finite State Machine and (security) process algebra [2] or based on logic induction of inductive logical programming paradigm [3]. Another approach is based on IDS software data modeling presented through "M2D2" example by Morin et al., [10].

In this contribution our approach is based on a formal description of state oriented dynamics by coalgebra, where a coalgebra (as formalized real IDS) is

Received by the editors: October 13, 2016.

2010 *Mathematics Subject Classification.* 18C10, 68M10.

1998 *CR Categories and Descriptors.* I.2.7 [**Mathematical Logic and Formal Languages**]: Formal Languages – *Language models*.

Key words and phrases. category theory, coalgebra, coalgebraic signature, IDS.

a pair consisting of a set called state space and coalgebraic structural function. This function represents the dynamics of a coalgebra. We define category of infinite packet stream, polynomial endofunctor over it and we model a IDS's behavior during possible network intrusion as a coalgebra for an appropriate polynomial endofunctor.

We see the benefits of this approach in the idea, that we construct formal model for observing program system behavior stepwise. Based on that, we can detect undesirable behavior and design verifiable model for the development of program systems. Such a model will guarantee program system's design correctness before its implementation. For example, the software engineering data modeling approach requires testing after any change in data model.

In our recent works [8],[13], [9], in the field of formal description of IDS behavior, we have captured network attacks in our real laboratory environment and described them using resource oriented logical system's behavioral formulæ which tasks are logical inference over the states of dynamic program system [14]. Here, we follow another structural point of view [12], where we model IDS's behavior through a coalgebra for polynomial endofunctor. In this contribution, we extend our whole work by investigating relations between coalgebraic (abstract) signatures and real network intrusion signatures.

2. Basic notions

In this section we introduce basic notions from the field of IDS, category theory and coalgebras. According to Rozenblum [15], the IDS is a software application or a hardware device which monitors computer system or its network enviroment. The IDS's main purpose is to protect computer network and prevent intrusions. It monitors all network traffic and based on the known patterns, it evaluates potential threats or network attacks. In our recent work [13], we have used the open source network intrusion detection system [15], called Snort.

Category theory was firstly founded in mathematics as a formalism for describing algebraic structures [1]. A category is a mathematical structure with well defined properties. Nowdays [7], category theory is a universal abstract frame used for description of various structures as mathematical, algebraic or abstract data structures used in computer science. Coalgebras are based on category theory in computer science [17] and can serve for modeling behavior of programs and program systems.

3. Modeling of the IDS behavior as a coalgebra for polynomial endofunctor

The aim of our approach in the field of a formal description of the IDS is modeling its behavior by coalgebra for polynomial endofunctor [6]. For formulating this approach it is necessary to follow the these four steps:

A. Definition of a many-typed coalgebraic signature containing the packet attributes and actions of the IDS based on possible intrusions.
B. Construction of the category of packets.
C. Specification of the polynomial endofunctor over constructed category.
D. Modeling of IDS as a coalgebra for polynomial endofunctor.

3.1. Many-typed coalgebraic signature.

According to Mihályi & Novitzká [7], many-typed coalgebraic signature is defined formally as an ordered pair:

$$\Sigma = (\mathcal{T}, \mathcal{F}), \tag{1}$$

where:

- $\mathcal{T}$ is a class of types names and
- $\mathcal{F}$ is a class of names of operations over types.

The fragment of many-typed coalgebraic signature containing structure of IDS packet handling, packet structure and its characteristic behavior is depicted in the Fig. 1.

```
BEGIN Signature
  $\Sigma_{packet}$
  Begin types
    $\mathcal{T} = \{actions, protocol, ipaddr, natip, nat\}$
  end
  Begin opns
    $\mathcal{F} = \{alert, log, pass, drop, reject :\rightarrow actions$
    $arp, icmp, tcp :\rightarrow protocol$
    $ttl :\rightarrow nat$
    $home\_port :\rightarrow nat$
    $exter\_net : natip \times natip \times natip \times natip \rightarrow ipaddr$
    $home\_net : natip \times natip \times natip \times natip \rightarrow ipaddr$
    $\ldots\}$
  end
END
```

Figure 1. Many-typed coalgebraic signature of packet structure

The type *actions* represents how IDS evaluates caught packet and its operational specifications denote packet structure description. Please note that the content of the signature fragment depicted in Fig. 1 is adjusted to the example in the section 4.3. We present description of the abbreviations used in the figures 1, 3, 4 in the following tables 1, 2.

Table 1. The signature type names description

Name	Description
actions	possible reactions of IDS on observed network activities
protocol	network protocol
ipaddr	IP address
natip	domain for IP addresses from $0 - 255$
nat	domain for natural number
$flow$	snort keyword for the traffic flow
tcp_flags	TCP flag bits

Table 2. The signature operational specification description

Name	Description
arp	address resolution protocol
icmp	Internet Control Message Protocol
tcp	Transmission Control Protocol
home_port	number of the local service
ttl	time to live of the packet
exter_net	external network IP address
home_net	local network IP address
redir_host	spoofed IP address
stateless	parameter of the snort keyword $flow$
$F, P, U, 12$	parameters of the snort keyword tcp_flags
icode	ICMP code
itype	ICMP type value

3.2. **Category of packets.** Barr and Wells in their work [1] define formally the category **C** as a mathematical structure consisting of

- class of objects $C_o = \{X, Y, Z...\}$ and
- class of morphisms $C_m = \{f, g, h...\}$.

The morphism $f \in C_m$ between two objects $X, Y \in C_o$ is denoted as $f : X \to Y$ (one can use a different notation: $X \xrightarrow{f} Y$), where X is domain and Y is codomain of the morphism f. Every object $X \in C_o$ has an identical morphism $id_X : X \to X$. The morphisms between categories, e.g. $F : \mathbf{C} \to \mathbf{D}$ are called functors.

Every category $\mathbf{C}$ possesses the most important property that the morphisms are composable: for every $X, Y, Z \in C_o$ and $f, g \in C_m$ where $f : X \to Y, g : Y \to Z$, there is a morphism $g \circ f : X \to Z$. The morphisms can be composed only if codomain of f and domain of g are the same.

Now, we construct the category of packets **Packets** which is defined as follows

- the objects are treated packets denoted as $p_1, p_2, p_3 ...$ and
- the morphisms $n : p_i \to p_{i+1}$, where $i \in \mathbb{N}$, are transitions between packets.

The model of the category **Packets** is depicted in the Fig. 2 bellow:

$$\begin{tikzcd}
p_1 \arrow[r, "n"] & p_2 \arrow[r, "n"] & p_3 \arrow[d, "n"] \\
\ldots & p_5 \arrow[l, "n"'] & p_4 \arrow[l, "n"']
\end{tikzcd}$$

FIGURE 2. Category **Packets**

Every object p possesses universal projecting property [7], defined as follows:

- $f : p \to ttl$;
- $g : p \to protocol$;
- $h : p \to home\_net$ and
- $i : p \to exter\_net$.

3.3. Specification of the polynomial endofunctor over category of packets.

Formally, an endofunctor F [17] is a functor which has the same domain and codomain. Kock in his work [4] defines a polynomial endofunctor T as an endofunctor which is constructed by the following polynomial operations

$$TX ::= X \mid X \times Y \mid X + Y \mid X^Y. \tag{2}$$

In this step we specify the polynomial endofunctor over constructed category **Packets** as

$$T : \mathbf{Packets} \rightarrow \mathbf{Packets}, \tag{3}$$

that is defined for objects and morphisms of the category **Packets** as follows

$$T(p) = X \times p \qquad \textit{and} \qquad T(n(p)) = X \times n(p), \tag{4}$$

respectively, where X denotes observable values from the particular packet stream [4], [5]. This could be observed by transitional coalgebraic structure

$$\langle \mathit{head}, \mathit{tail} \rangle : \rho_p \rightarrow T(\rho_p), \tag{5}$$

where

- the operation *head* returns the first packet of packet stream and the operation *tail* returns the rest of packets of the packet stream, and they are defined as follows
  - $\mathit{head} : \rho_p \rightarrow p$;
  - $\mathit{tail} : \rho_p \rightarrow \rho_p$;
- ρ_p denotes a stream of packets, whereby it is a composition of morphisms in the category **Packets**:

$$p_1 \xrightarrow{n} p_2 \xrightarrow{n} p_3 \xrightarrow{n} \dots \tag{6}$$

Current observation of a packet p of a stream of packets dynamic p_1, p_2, p_3... can be expressed as follows

$$\mathit{head}(\rho_p), \mathit{head}(\mathit{tail}(\rho_p)), \mathit{head}(\mathit{tail}(\mathit{tail}(\rho_p)))... \tag{7}$$

i.e.

$$p_1 = \mathit{head}(\rho_p),\ p_2 = \mathit{head}(\mathit{tail}(\rho_p))... \tag{8}$$

3.4. **Coalgebra as an IDS.** Let us consider category $\mathbf{C}$ and a polynomial endofunctor T over category $\mathbf{C}$. Then according to Moss [11], coalgebra for a polynomial endofunctor is defined formally as an ordered pair

$$(X, \omega), \tag{9}$$

where

- X called the state space of the objects of category $\mathbf{C}$ and
- ω is the coalgebraic structure [16] as morphisms of category $\mathbf{C}$ such as: $\omega : X \rightarrow TX$.

Now based on the introduced definition of polynomial endofunctor (3) in the subsection (3.3), we coalgebraically model the IDS as a coalgebra for polynomial endofunctor T, i.e. T-coalgebra for infinite packet stream ρ_p as follows

$$(\rho_p, \langle head, tail \rangle), \tag{10}$$

where ρ_p is state space and $\langle head, tail \rangle$ is pair of functions as structural function of the T-coalgebra.

4. Network intrusion signatures as coalgebraic signatures

Based on our previous work [12], we demonstrate our approach on two types of intrusions:

- The NMAP network intrusion i.e. combined portscan using the network mapper tool.
- The ARP Spoofing network attack i.e. network attack caused by arpspoof network tool that "redirects packets from a target host on the LAN intended for another host on the LAN by forging ARP replies" [18].

In this section, we present how the real network signatures can be coalgebraically specified. Firstly we present the Snort's signatures, then we abstract important equations (conditions which have to be fulfilled for detection of the intrusion), we depict them in the form of tables and then we coalgebraically specify the signatures.

4.1. Signature of the NMAP network intrusion.

The Snort signature for `Scan NMAP XMASS` rule defined by Snort identification number 1228 is defined as follows:

```
alert tcp EXTERNAL_NET any -> HOME_NET any
(msg:"SCAN nmap XMAS"; flow:stateless;
flags:FPU,12; reference:arachnids,30;
classtype:attempted-recon; sid:1228; rev:7;)
```

The corresponding equation table is depicted in Table 3.

Table 3. Values for NMAP intrusion symptoms

$exter\_net$	=	any
$home\_net$	=	any
$home\_port$	=	7
$protocol$	=	tcp
$flow$	=	$stateless$
tcp_flags	=	$F, P, U, 12$

The corresponding coalgebraic signature is depicted in Figure 3.

```
BEGIN Signature
    $\Sigma_{nmap}$
    Begin types
        $\mathcal{T} = \{action, flow, ipaddr, nat, natip,$
        $protocol, tcp\_flags\}$
    end
    Begin opns
        $\mathcal{F} = \{alert :\to action$
        $tcp :\to protocol$
        $F, P, U, 12 :\to tcp\_flags$
        $home\_port :\to nat$
        $stateless :\to flow$
        $exter\_net : natip \times natip \times natip \times natip \to ipaddr$
        $home\_net : natip \times natip \times natip \times natip \to ipaddr\}$
    end
END
```

Figure 3. Many-typed coalgebraic signature of NMAP network intrusion

4.2. **Signature of the ARP Spoofing network attack.** The Snort `ARP Spoofing` signature for `ICMP redirect host` rule with Snort's ID 472 is defined as follows:

```
alert icmp $EXTERNAL_NET any -> $HOME_NET any
(msg:"ICMP redirect host"; icode:1; itype:5;
reference:arachnids,135; reference:cve,
1999-0265; classtype:bad-unknown; sid:472;
```

`rev:4;)`

The corresponding equation table is depicted in Table 4.

TABLE 4. Values for ARP spoofing intrusion symptoms

exter_net	=	*any*
home_net	=	*any*
redir_host	=	*any*
protocol	=	*icmp*
icode	=	1
itype	=	5

The corresponding coalgebraic signature is depicted in Figure 4.

```
BEGIN Signature
    $\Sigma_{aSpoof}$
    Begin types
        $\mathcal{T} = \{action, flow, ipaddr, nat, natip,$
        $protocol\}$
    end
    Begin opns
        $\mathcal{F} = \{alert :\to action$
        $icmp :\to protocol$
        $icode, itype :\to nat$
        $stateless :\to flow$
        $exter\_net : natip \times natip \times natip \times natip \to ipaddr$
        $home\_net : natip \times natip \times natip \times natip \to ipaddr$
        $redir\_host : natip \times natip \times natip \times natip \to ipaddr\}$
    end
END
```

FIGURE 4. Many-typed coalgebraic signature of ARP spoofing network intrusion

4.3. **Example of IDS behavior modeled as coalgebra for polynomial endofunctor T.** The coalgebra 10 observes the packet stream ρ_p and it checks every packet for specified symptoms of the attack $\circ$. After that, it will responds by performing one of the actions defined in many-typed coalgebraic signature Σ_p depicted in the Fig. 1, by performing specific operation.

In this model example, we consider only two types of intrusions A, B, where A means NMAP network intrusion and B denotes ARP Spoofing network attack.

The actual content of intrusions A, B and the symptoms are presented in the tables (3), (4) respectively. The first column here shows operational specifications of particular symptoms in order to coalgebraic signatures depicted in Fig. 3 and Fig. 4. The second one represents values which are characteristic for a real network intrusion behavior based on Snort's signature mentioned below under the Snort's ID numbers 1228 and 472 respectively. In this manner we are able to model every known type of the attacks with this method.

Behavior of the IDS "by steps" is expressed by the sequence:

$$\begin{array}{rcll}
(p_1, p_2, p_3, p_4, ...) & \mapsto & (p_1, (p_2, p_3, p_4), \epsilon \mapsto pass) & \mapsto \\
 & \mapsto & (p_1, p_2, (p_3, p_4), A \mapsto alert) & \mapsto \\
 & \mapsto & (p_1, p_2, p_3, (p_4), B \mapsto alert) & \mapsto \\
 & \mapsto & ... &
\end{array}$$

In this example we can observe the sample situation when the coalgebra for polynomial endofunctor provides sequence of four packets in a network stream. It does not detect a match with the symptoms from the tables (3) and (4) in the packet p_1, therefore it performs the operation *"pass"* defined in the signature depicted in the Fig.1 (we use the letter ϵ for denoting that coalgebra has found no match with attack symptoms). In the packet p_2 it recognizes the network intrusion `SCAN nmap XMAS` and it performs the operation *"alert"*. At the packet p_3 it captures the `ICMP redirect host` attack and it also performs the operation *"alert"*.

5. Ackwnowledgments

This work was supported by the Faculty of Electrical Engineering and Informatics, Technical University of Košice under the contract No. FEI-2015-18: Coalgebraic Models of Component Systems.

6. Conclusion

Main goal of this paper was to demonstrate our coalgebraic approach of formal description of program systems. We have showed how IDS's behavior can be formally described by coalgebra for a polynomial endofunctor and we have modeled its behavior when it detects specific network intrusions. Here we have worked only with few specific network intrusions, but it is possible to model any known network intrusion with this approach. Based on that, we can create an verifiable construction model of observable behavior, which allows as for example to remove undesirable behavior of program system. In this contribution we have shown small but important part of our bigger proposition of verifiable model for IDS design.

In the future, we would like to extend our approach and we would like to do more specific formal description of program systems and programming paradigms such as component based programming.

References

[1] M. Barr, CH. Wells. *Category theory for computing science*, Prentice Hall International (UK) Ltd., 66 Wood Lane End, Hertfordshire, UK, (1998).

[2] A. Durante, R. Di Pietro, L. V. Mancini, *Formal specification for fast automatic ids training*, Formal Aspects of Security, First International Conference, FASec 2002, Springer Berlin Heidelberg, (2003), pp. 191-204.

[3] C. Ko, *Logic induction of valid behavior specifications for intrusion detection*, Security and Privacy, S&P 2000. Proceedings. 2000 IEEE Symposium on. IEEE, (2000), pp. 142-153.

[4] J. Kock, *Notes on Polynomial Functors*, Departament de Matematiques, Universitat Autonoma de Barcelona, Barcelona, Spain, (2009).

[5] A. Kurz, *Coalgebras and Modal Logic*, CWI, Amsterdam, Netherlands, (2001).

[6] D. Mihályi, V. Novitzká, M. Ľaľová, *Intrusion Detection System Episteme*, Central European Journal of Computer Science. Vol. 2, no. 3, (2012), pp. 214-220.

[7] D. Mihályi, V. Novitzká, *Princípy duality medzi konštruovaním a správaním programov*, Equilibria, Košice, (2010).

[8] D. Mihályi, V. Novitzká, *Towards to the Knowledge in Coalgebraic model IDS*, Computing and Informatics, 33, 1, (2014), pp. 61-78.

[9] D. Mihályi, V. Novitzká, P. Pražák, P. Popovec, *Network routing modelled by game semantics*, Studia Universitatis Babes-Bolyai, Informatica 57, no. 4 (2012).

[10] B. Morin, et al., *M2D2: A formal data model for IDS alert correlation*, Recent Advances in Intrusion Detection, 5th International Symposium, RAID 2002, Springer Berlin Heidelberg, (2002), pp. 115-137.

[11] L. Moss, *Coalgebraic logic*, Annals of Pure and Applied Logic, Volume 99, Issues 13, Department of Mathematics, Indiana University, Bloomington, pp. 241-259, USA, (1997).

[12] J. Perháč, D. Mihályi, *Coalgebraic modeling of IDS behavior*, Informatics 2015, Danvers, IEEE, (2015), pp. 201-205.

[13] J. Perháč, D. Mihályi, *Intrusion Detection System Behavior as Resource-Oriented Formula*, Acta Electrotechnica et Informatica, Vol. 15, no. 3, (2015), pp. 9-13.

[14] J. Perháč, D. Mihályi, V. Novitzká, *Between syntax and semantics of resource oriented logic for IDS behavior description.* Journal of Applied Mathematics and Computational Mechanics, Volume 15, Issue 2, (2016), pp. 105-118.

[15] D. Rozenblum, *Understanding Intrusion Detection Systems*, SANS Institute InfoSec Reading Room, (2001).

[16] V. Slodičák, P. Macko, *How to apply linear logic in coalgebraical approach of computing*, CECIIS 2011, Proceedings of the 22nd Central European Conference on Information and Intelligent Systems, September 21st-23rd 2011, Varaždin, Croatia. - Varaždin, University of Zagreb, (2011), pp. 373-380.

[17] V. Slodičák, P. Macko, *Some New Approaches in Functional Programming Using Algebras and Coalgebras*, Electronic Notes in Theoretical Computer Science. Vol. 279, no. 3, (2011), pp. 41-62.

[18] D. Song. Arpspoof manual page, Online: `http://code.tools/man/8/arpspoof/`.

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Košice, Slovak Republic

E-mail address: {`Jan.Perhac, Daniel.Mihalyi`}`@tuke.sk`