A Coalgebra as an Intrusion Detection System

Daniel Mihályi, Valerie Novitzká

Department of Computers and Informatics
Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9, 042 00 Košice, Slovakia
valerie.novitzka@tuke.sk, daniel.mihalyi@tuke.sk

Abstract: In this paper we construct a coalgebra for an intrusion detection system to describe the behaviour of a packet stream together with selected actions in the case of intrusions. We start with an extension of the notion of the many-typed signature to the generalised signature and we construct the category of packets as a basic structure of our approach. A defined endofunctor captures the expected behaviour of the packet stream. The constructed coalgebra enables the description of the behaviour of the packet stream together with the reaction to intrusions.

Keywords: Coalgebra; Category theory; Intrusion detection system

1 Introduction

The main purpose of our research [5], [6], [7] is the construction of behavioural categorical models based on coalgebras for large program systems. There are only quite simple examples of using coalgebras in actual programs. In this contribution we show how it is possible to use our results for nontrivial systems from the area of real applications in informatics. We chose an Intrusion Detection System (IDS) to show how its behaviour can be modelled in categorical manner by a coalgebra.

The main purpose of an IDS is to disclose potential unwanted network activities. Many contemporary tendencies and trends are mostly pointed towards signature-based methods for attack-recognition. The idea of this method rests on the comparison of actually observed network traffic and the collection of known attack descriptions [4]. In our approach, we present another abstract means of notion *signature*. The well known notion of the universal algebra, a *many-typed signature* we extend to a *generalised signature*. Because there we deal with complex packet structures, we need to describe them in more complex mathematical structures. We use families of sets to describe heterogeneous informatic structures, e. g. records, and enclose them into a category.

In our approach, we formulate an IDS in the theory of coalgebras of semipolynomial endofunctors [3] over generalised signatures which are depicted in an abstract frame of category theory [1]. Our approach we formulate in the following steps:

1 first of all, we define a generalised signature containing the structure of treating packets and its chosen properties;

2 in the next step we construct the category of packets;

3 then we determine the semipolynomial endofunctor over this category;

4 afterward we characterize symptoms of network attacks and intrusions;

5 finally we excogitate the coalgebra of a semipolynomial endofunctor over a category of packets by means of which we describe the behaviour of infinite packet streams.

2 Generalised Signature

First of all we have to construct a *generalised signature* as an extension of the algebraic signature as a pair

$$\sum_p = (\bar{T}, F) \tag{1}$$

consisting of a finite collection of *Church's type names* $\bar{T}$ and a finite collection of *operation specifications* on Church's types denoted by *F*. This set includes the structure specification of treating packets e.g. *version*, *ttl*, *protocol,* etc. and structural properties of packets like *dsize*, *itype*, *content* etc. In operation specifications we distinguish three families:

1 *constructor-operation specifications* denoted by $f: \sigma_1 \lozenge \sigma_2 \rightarrow \tau;$

2 *deconstructor-operation specifications* denoted by $f: \sigma \rightarrow \tau_1 \lozenge \tau_2;$

3 *derived operation specifications* denoted by $f: \sigma \rightarrow \tau$ and $f: \sigma_1 \lozenge \sigma_2 \rightarrow \tau_1 \lozenge \tau_2.$

where $\sigma, \tau \in \bar{T}$ are arbitrary types from $\bar{T}$. The symbol $\lozenge$ is a placeholder for the type operation of product, coproduct and function. Then the specification of the signature Σ_p is treating a packet which we denote by *p*.

Table 1
IDS Signature

BEGIN Signature
Σ_p
Begin types

$\bar{T}$ ={*actions, protocols, ips, port, message, natip, nat0, char*}
End types
Begin opns
$\mathcal{F}$ = {*alert, log, drop, activate:* → *actions,*
255, icmp, tcp: → *protocols,*
ttl: → *nat0,*
port: → *nat0,*
mac addr: hex x *hex* x *hex* x *hex* x *hex* x *hex* → *mac,*
ip addr: natip x *natip* x *natip* x *natip* → *ips,*
ver : → *nat,*
message : → *char,*
dsize : → *nat0,*
content : → *char,*
itype : → *nat0}*
End opns
END Signature

3 Category *Packet*

In the second step we need to construct the category $Packet$ (Figure 1) of packets, where objects are treated packets denoted by $p_1, p_2, \ldots$ as non trivial heterogeneous structures – records, and morphisms *next:* $p_i \rightarrow p_{i+1}$ express homomorphous transition into the next packet of a given stream.

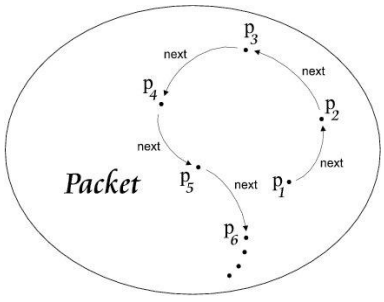

Figure 1
The category of packets

For any object p holds the universal mapping property mentioned in [1] in the following way: for any object $p \in Packet_{Obj}$ and projection morphisms $f:p \rightarrow ver$, $g:p \rightarrow ttl$, $h:p \rightarrow protocol$, $i:p \rightarrow s\_addr$, $i:p \rightarrow d\_addr$ there exists one (multiple) morphism

$$\langle f, g, h, i, j \rangle : p \rightarrow Nat \times Nat0 \times Protocols \times IPs \times IPs \tag{2}$$

depicted in the Figure 2 by dashed arrow.

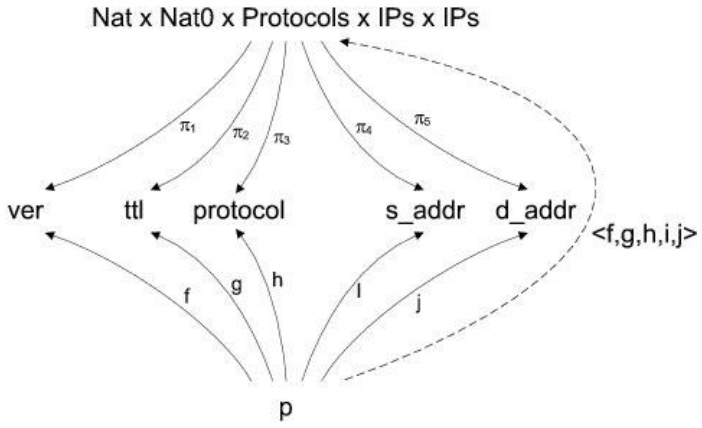

Figure 2
The universal projection property on structure p

3.1 Stream Automata

With respect to problems related to intrusion detections, we start from the theory of *stream automata* published in [2]. The authors represent trivial models of dynamical systems behaviour on infinite streams consisting of set elements. For instance, we can define an automata as a triple

$$SA = (Q, hd : Q \rightarrow P, tl : Q \rightarrow Q) \tag{3}$$

where Q is a set of (internal) states, $hd: Q \rightarrow P$, resp. $tl: Q \rightarrow Q$ are *head* resp. *tail* functions of a given stream.

If we consider trivial packets we also get a "trivial system" that can be described by display and one button. Then we can enunciate the principle: display packet when the button is pressed.

3.2 Coalgebra without Detection

In our approach, for a given trivial stream of packets without intrusive detection, we introduce the appropriate coalgebra *(ρ_p <hd,tl>)* in the following way. Infinite stream of packets we denote by ρ_p as state space of the coalgebraic structure

$$\langle hd, tl \rangle : \rho_p \to p \times \rho_p \tag{4}$$

We specify stream coalgebraic operations head resp. tail as *hd:* $\rho_p \to p$, resp. *tl:* $\rho_p \to \rho_p$ where ρ_p represents morphism compositions in the category $Packet$

$$p_1 \xrightarrow{next} p_2 \xrightarrow{next} \dots \tag{5}$$

We can formulate dynamics (behaviour) of infinite stream ρ_p as a sequence

$$\left(hd(\rho_p), hd(tl(\rho_p)), hd(tl^2(\rho_p)), \dots\right) \tag{6}$$

where $p_1 = hd(\rho_p)$, $p_2 = hd(tl(\rho_p))$,…

4 Semipolynomial Endofunctor

Next, we construct a semipolynomial functor over objects and morphisms of the category $Packet$ as

$$T : Packet \to Packet \tag{7}$$

defined in the following way

$$T(p) = X \times p \tag{8}$$

and

$$T(next(p)) = X \times next(p) \tag{9}$$

where X denotes observed values of a given packet. Then, the transition coalgebraic structure has the following form

$$\langle hd, tl \rangle : \rho_p \to T(\rho_p) \tag{10}$$

This structure gives us some observations of the network behaviour from outside based on observable values.

5 The Coalgebra

5.1 The Coalgebra with Detection

Now we extend the coalgebra introduced in 4.2 to the coalgebra with detection of unwanted network intrusions.

For the demonstration example, we show in Table 2 three selected specifications *A,B,C* of usual network intrusions by [8], whereas their real intendment is in parenthesis. We can consider the values listed below in the form of equalities as the symptoms of a potential network remote attack.

If from captured packet are observed some known symptoms mentioned above, then the coalgebra (system) responds by making one of the following preferred reactions, such as

• *alert,* which generates appropriate attention on the screen,

• *log,* for intrusion protocolling,

• *drop,* which ignores the intrusive fact by throwing away the incriminated packet and activation.

Table 2
Specifications of network intrusions

A	*B*	*C*
(ICMP Ping NMAP)	*(TCP Portscan)*	*(DOS Cisco attempt)*
IP Protocol == icmp	*MAC Addr == MACDAD*	*Port == 80*
dsize == 0	*IP Protocol == 255*	*dsize == 1*
itype == 8	*IP TTL == 0*	*content == "|13|"*

Now we need to extend the definition of the semipolynomial functor to include the detection of the known intrusions. We can formalize the activity of the whole system by mapping

$$atack(p) \mapsto (p, next(p), intrusion\_type(p)) \tag{11}$$

where *intrusion_type(p)* is a function of the form

$$intrusion\_type(p) : I \to actions \tag{12}$$

where *I* is a particular type of intrusion.

5.2 The Coalgebra as an IDS

As the last step, we construct coalgebra as intrusion detection system

$$\left(\rho_p, \left\langle hd, tl, intrusion\_type \right\rangle\right) \tag{13}$$

which is explicitly characterised by the following operations

- Immediate observation of treating packet *hd:* $\rho_p \rightarrow p$
- State modification *tl:* $\rho_p \rightarrow \rho_p$ and
- Generation of appropriate action (*intrusion_type(p):* $\rho_p \rightarrow p \circledR actions^I$) in the form

$$\left\langle hd, tl, intrusion\_type \right\rangle : \rho_p \rightarrow p \times \rho_p \times p \circledR actions^I \tag{14}$$

where $p \circledR actions^I$ expresses the generation of the appropriate reaction *actions* according to the given intrusion type $I=A+B+C$ in the appurtenant field of packet p, i.e. coincidence was captured between an intrusive pattern of network traffic and symptoms from Table 2.

5.2.1 Example

Behaviour of the system described by the coalgebra (13) can be modelled "step by step" by the following sequence

$$\begin{aligned} (p_1, p_2, \ldots) \quad &\mapsto \quad (p_1, (p_2, p_3, p_4,), A \mapsto alert) \quad \mapsto \\ &\mapsto \quad (p_1, p_2, (p_3, p_4,), \varepsilon) \quad \mapsto \\ &\mapsto \quad (p_1, p_2, p_3, (p_4,), C \mapsto alert) \quad \mapsto \\ &\mapsto \quad \ldots \end{aligned}$$

In the event that one of intrusions A,B or C is detected, some of the predefined actions from signature Σ_p are performed.

The example shows the situation where on any pattern of network traffic are treated packets p_1, p_2, p_3, p_4. On the packet p_1 was captured intrusion "*ICMP Ping NMAP*" by the specification A from Table 2, and on the packet p_2 was captured intrusion "*DOS Cisco attempt*" by the specification C from the same table.

5.3 The Final Coalgebra

Finally we turn our attention to constructing the final coalgebra. Let $\mathcal{T}coalg$ be the category of coalgebras over the semipolynomial endofunctor T where objects are

coalgebras on infinite data structures and morphisms are structure preserved homomorphisms between coalgebras. Its final object is the final coalgebra

$$\left(\rho_w, \left\langle observer, nextstat, i_t \right\rangle\right) \tag{15}$$

over the semipolynomial endofunctor T where *observer* is the generalized operation for performing an immediate observation on a data element of infinite data structure, *nextstat* is the next state operation and i_t is generator of the appropriate action.

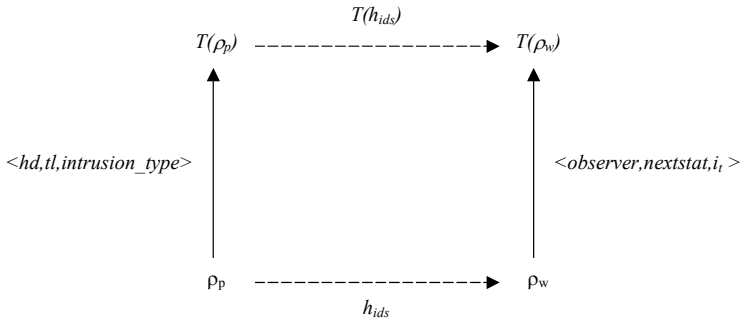

Figure 3
Homomorphism of the final coalgebra

For every operation *hd,tl* or *intrusion_type* of the intrusion detection coalgebra $(\rho_p, \langle hd,tl,intrusion\_type\rangle)$ in the packet state space ρ_p of the category *Packet* there exists a unique morphism (behavioural relation) in the category of coalgebras $Tkoalg$

$$\left\langle hd, tl, intrusion\_type \right\rangle \rightarrow \left\langle observer, nextstat, i_t \right\rangle \tag{16}$$

Where the diagram at Figure 3 commutes.

We call the homomorphism h_{ids}: $\rho_p \rightarrow \rho_w$ *infinite stream packet behavior* of a given computer network. This behaviour is realized stepwise by repeated evaluation of the coalgebraic structure. From these facts we see that the mapping h_{ids} captures stepwise particular packet observations by means of operation *hd*, which originate from the increased application of operation *tl*.

Conclusions

In this paper we have shown how coalgebras can be used for the modelling of real program systems. Our contribution contains the step by step construction of a coalgebra for an IDS. The constructed coalgebra describes the behaviour of infinite stream of packets with the detection of possible intrusions. This model covers also actions executed in the case of intrusions.

Our results demonstrate that coalgebras can be useful for a wide spectrum of large program systems. Of course, this paper deals only with one area of program systems, but in following research we will concern ourselves with the modelling of other systems, e.g. database systems or distributed systems by coalgebras.

In the future we would like to extend coalgebraic models with resource-oriented modal logics for proving bisimilarities on states produced by a system.

Acknowledgement

This work was supported by VEGA Grant 1/0175/08: Behavioural Categorical Models for Complex Program Systems.

References

[1] Barr, M., Wells, C.: Category Theory for Computing Science. Prentice Hall International (UK) Ltd., 66 Wood Lane End, Hertfordshire, UK, 1990

[2] Hasuo, I.: Modal Logics for Coalgebras-A Survey. Tech. rep., Institute of Technology, Tokyo, 2003

[3] Jacobs, B.: Introduction to Coalgebra. Towards Mathematics of States and Observations (draft), 2006

[4] Kazachkin, D. S., Gamayunov, D. Y.: Network Traffic Analysis Optimization for Signature-based Intrusion Detection Systems. Computational systems lab of Moscow State University's Faculty of Computational Math and Cybernetics, 2008

[5] Novitzká Valerie, Jenčík Marián, Mihályi Daniel, Slodičák Viliam, Ľaľová Martina: Behaviour of Program Systems in Terms of Categories, Computer Science and Technology Research Survey, Košice, Elfa, 2009, pp. 31-36, ISBN 978-80-8086-131-5

[6] Novitzká, V., Mihályi, D., Verbová, A. Coalgebras as Models of System's Behaviour. In AEI 2008, International Conference on Applied Electrical Engineering and Informatics '2008 (Athens, Greece, 2008), DCI FEI Technical University, Košice, pp. 31-36

[7] Slodičák Viliam, Mihályi Daniel: Coalgebras for Program Behavior in Toposes and Comonads, Proceedings of the Tenth International Conference on Informatics - Informatics 2009, Košice, Herľany, November 23-25, 2009, Košice, elfa, s.r.o., 2009, 10, pp. 125-135, ISBN 978-80-8086-126-1

[8] The Snort Team, s.-t. Snortusers manual. The Snort Project, 2008