The Research of Formal Specification of Intrusion Detection Rules

Sun Meifeng Gong Jian
Department of Computer Science and Technology
Southeast University
Nanjing 210096, P.R. China
{msun, jgong}@njnet.edu.cn

Abstract

This paper first presents a taxonomy of detection language and then a methodology for estimating detection languages in term of their expressive ability, succinctness and detection complexity is developed.

1. Introduction

Many detection languages have been proposed to specify attack signatures in order to detect misuses. An interesting question thus arises: How does one evaluate and compare different formalisms of attack signatures? Besides the practical aspect of this question with respect to choosing the "best" formalism for a given application, environment, and resource constraints, a methodology for comparing and evaluating attack signature representation methods may lead to useful introspection, new insights, and to the discovery of better approaches.

This paper first introduces a kind of taxonomy to classify detection languages in section 2. Then section 3 formally defines three metrics: expressibility, succinctness and detection complexity and compares various formalisms of attacks by these metrics.

2. The Classification of Detection Languages

First, let us explain two operators. Operator ° concatenates two sequences of filter. Operator • means the arbitrary concatenation of two sequences of filter, regardless of the order. For example, ω1•ω2 = {'ABCD','ACBD','ACDB','CABD','CADB','CDBA'}. Let S1, S2, S be arbitrary scenarios, where S is combined by S1 and S2:

Definition 1:

then (S = S1 then S2): **S** = {ω1°ω2: ω1∈S1,ω2∈S2};

or (S = S1 or S2): **S** = S1∪S2;

and (S = S1 and S2): **S** = {ω1•ω2: ω1∈S1, ω2∈S2}.

Definition 2:

γ Relation: Let M' be the set of detection languages, γ = {(M_1,M_2)： M_1,M_2∈M' and ∀x： x∈{then,and,or}, "M_1 supports x" ⇔"M_2 supports x"}

Obviously, γ is reflective, symmetrical, and transitive.

Definition 3:

Classification μ: **μ** is a classification on M' based on γ relation.

μ = {[{}], [{then}] [{and}], [{or}], [{then , and}], [{then , or}], [{and , or}] ， [{then , and, or}]}，where ∀x: x⊆{then, and, or}, **[x]** = { M | M ∈ M' and M supports and only supports the combination ways in x}. For convenience, we omit the set notation of x in [x], that is, [{then,or}] is written as [then,or]. Colored Petri-net[1] is an element of [then,or ,and], however

Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks (LCN'04)
0742-1303/04 $ 20.00 IEEE

statl[2] belongs to [then,or]

The classification μ is based on the syntax elements of detection languages, which guarantees non-ambiguity and repeatability. Moreover, μ is mutual exclusive and complete, because it is the classification based on equivalence relation.

3. Some Metrics of Estimating Detection Languages

As any attack Scenario can be described by enumerating all filter sequences, which is the most verbose form of description, it becomes a baseline for estimating succinctness. The signature base generated by enumeration is thus called the baseline of signature base $\Re$, written as KL($\Re$).

Let M' be the set of detection language, $\Re$' be the set of signature base $\Re$, **DL(x)** returns signature base x's detection language.

Definition 4: Relative Expressibility

The Equivalence of Expressibility LG=:

$LG= = \{(A, B): A, B \in M' \wedge \forall x(x \in \Re' \wedge DL(x)=A \rightarrow \exists y(y \in \Re' \wedge DL(y)=B \wedge Kl(x)=Kl(y))) \wedge \forall y(y \in \Re' \wedge DL(y)=B \rightarrow \exists x(x \in \Re' \wedge DL(x)=A \wedge Kl(x)=Kl(y)))\}$

Similarly, we can define LG< and LG>.

Observation 1: 'then' is the prerequisite operator in describing multiphase scenarios. The expressibility will not be enhanced even if more ways of combination are supported.

Conclusion 1: [ ] LG< [then] LG= [then,or] LG= [then,and] LG= [then,or,and]

Definition 5: Succinctness

f: M' → N is a function from the set of detection languages to natural number. $\forall \Re$：DL($\Re$)=M , f(M) is defined as the maximum growth rate of size when $\Re$ is transformed into Baseline($\Re$). The greater f(M) is, the more succinct M is.

Observation 2: Under the condition that 'then' has been supported, the succinctness of expression will be enhanced in polynomial if 'or' is supported as a plus, and in exponential if 'and' is supported as a plus.

Conclusion 2:

[then]<[then,or]<[then,and] <[then,or,and]

Definition 6: Detection Complexity

$\mathfrak{I}$: M'→ N is a function from the set of M to the set of natural number. $\forall \Re$：DL($\Re$)=M, $\mathfrak{I}$(M) is the maximum growth rate of reasoning time when $\Re$ is transformed into Baseline($\Re$).$\mathfrak{I}$(M) is a function of the size of signature base. The greater $\mathfrak{I}$(M) is, the more stronger M's detection ability is.

Observation 3: Under the condition that 'then' operator has been supported, the detection complexity will be increased in polynomial if 'or' operator is supported as a plus, and in exponential if 'and' operator is supported as a plus.

Conclusion 3: [then] < [then,or] < [then,and] < [then,or,then]

4. Conclusion

From the perspective of expressive ability, detection languages can be categorized into two classes, that is, unit-event languages and multiphase languages. The multiphase languages have the same expressive ability, though they are of different forms. The combination method 'then' is the prerequisite for describing multiphase attack. The combination method 'and' and 'or' can enhance the succinctness and detection complexity of languages

REFERENCES

[1] S. Kumar. Classification and Detection of Computer Intrusions[PhD thesis], Dept. of Computer Science, Purdue University, USA, 1995.

[2] G.Vigna, S.T. Echmann, and R.A. Kemmerer. STATL: An Attack Language for State-based Intrusion Detection. Dept of Computer Science University of California Santa Barbara, 2000.