

# A Temporal Logic Based Framework for Intrusion Detection

Zanolin Lorenzo

September 5, 2023

## Abstract

The purpose of this paper is to introduce MONID, which is a framework created for system intrusion detection. This framework uses EAGLE, a rich and effectively monitorable logic, to express intrusion patterns using temporal logic formulas; EAGLE's ability to include data values and parameterized recursive equations makes it possible to represent security threats that include complex temporal event sequences and attacks with intrinsically statistical signatures succinctly. This tool can be used in off-line and real-time scenarios. The implementation uses an algorithm for online monitoring that matches descriptions of the lack of an assault with indications of system execution; an alarm is set off whenever the standard is broken.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>EAGLE, a Temporal Monitoring Logic</b>	<b>2</b>
2.1	Basics . . . . .	3
2.2	Syntax and Semantics . . . . .	4
2.3	Evaluation algorithm . . . . .	5
2.4	Relationship to LTL+P . . . . .	7
<b>3</b>	<b>MONID: Structure analysis</b>	<b>8</b>
<b>4</b>	<b>Example of attacks</b>	<b>9</b>
<b>5</b>	<b>Conclusions</b>	<b>9</b>

## 1 Introduction

Even with all of the advances in computer security research, totally safe computer systems remain a long way off. Almost every large and complicated computer system nowadays contains vulnerabilities. Intrusion detection means maintaining constant surveillance on a system in order to detect any misuse of these weak areas as soon as feasible so that they can be repaired.

There are three Intrusion Detection System approaches, according to the literature: *signature-based*[4], *anomaly-based*[6] and *hybrid*[5]. The first approach aims to identify patterns and match them with known signs of intrusions relying on a database of previous intrusions. If activity within the network matches the “signature” of an attack or breach from the database, the detection system creates an alert. This approach has a low false-alarm rate, but it requires us to know the patterns of security attacks in advance and previously unknown attacks would go undetected. In contrast, *anomaly-based* is capable to detect new attacks since an alarm is raised if an observed behavior deviates significantly from pre-learned normal behavior. Finally, a *hybrid system* combines the best of both worlds by looking at patterns and one-off events, a Hybrid Intrusion Detection system can flag new and existing intrusion strategies.

In this paper we will focus on *signature-based* approach using temporal logic; we use EAGLE[3, 1] to specify a system’s attack-safe behavior. EAGLE allows recursively built temporal formulas, parameterizable by both logical formulas and data expressions, across three primitive modalities: “next”, “previous”, and “concatenation”. The logic allows us to describe temporal patterns involving reasoning about data-values observed in individual events. Unlike LTL, EAGLE allows us to design attacks with fundamentally statistical characteristics; password guessing attacks and ICMP-flood denial of service attacks are two examples. We will monitor EAGLE formulas using an online algorithm that processes each event as it happens and updates the monitored formula to store a relevant summary. An intrusion alarm is triggered if, for any reason, the updated formula turns out to be false; as a result, the entire process operates in real-time. We’ll see how the implementation, MONID, examines the event stream to see if the monitored formula (which is used to hold the pertinent summary of the system) is being broken.

In Section 2 we will analyze the basics of EAGLE. In Section 3 we will analyze MONID framework and some evaluations on the performances. In Section 4 we will see some attacks specifications. In section 5 we will conclude.

## 2 EAGLE, a Temporal Monitoring Logic

According to [2], EAGLE offers a succinct but powerful set of primitives, essentially supporting recursive parameterized equations, with a minimal/maximal fix-point semantics together with three temporal operators: next-time ( $\bigcirc$ ), previous-time ( $\odot$ ), and concatenation ( $\cdot$ ).

## 2.1 Basics

In EAGLE recursion definitions are supported; as example, in the current framework we can build the following definitions:

$$\begin{aligned}\underline{\min} \text{ Next}(\underline{\text{Form}} F) &= \bigcirc F \\ \underline{\max} \text{ Always}(\underline{\text{Form}} F) &= F \wedge \bigcirc \text{Always}(F) \\ \underline{\min} \text{ Eventually}(\underline{\text{Form}} F) &= F \vee \bigcirc \text{Eventually}(F) \\ \underline{\min} \text{ Until}(\underline{\text{Form}} F_1, \underline{\text{Form}} F_2) &= F_2 \vee (F_1 \wedge \bigcirc \text{Until}(F_1, F_2))\end{aligned}$$

As we can see, rules are parameterized by an EAGLE formula (of type  $\underline{\text{Form}}$ ), which means that we will be able to write EAGLE formulas such as  $\text{Always}(\text{Eventually}(x > 0))$ . Also, the Always operator is defined as maximal solution of the equation  $X = F \wedge \bigcirc X$ , while the Eventually operator represents the minimal solution to the equation  $X = F \vee \bigcirc X$ .

Assume that we want to state the following property: "Whenever there is a login by any user  $x$ , then eventually the user  $x$  logs out". In LTL we can write the following formula:

$$\Box((\text{action} = \text{login}) \rightarrow \underline{\text{let}} k = \text{userId} \underline{\text{in}} \Diamond(\text{action} = \text{logout} \wedge \text{userId} = k))$$

In this formula we use the operator  $\underline{\text{let}} \_ \underline{\text{in}} \_$  to bind the value of  $\text{userId}$  in the current event to the local variable  $k$  whenever  $(\text{action} = \text{login})$  in the current event; we then impose the condition that the value of  $\text{userId}$  in some event in future must be same as the user id bound to  $k$  and that the action of the event must be logout. In EAGLE we can express the same property with the following rules:

$$\begin{aligned}\underline{\min} \text{ EvLogout}(\underline{\text{string}} k) &= (\text{action} = \text{logout} \wedge \text{userId} = k) \vee \bigcirc \text{EvLogout}(k) \\ \underline{\text{mon}} M_2 &= ((\text{action} = \text{login}) \rightarrow \text{EvLogout}(\text{userId}))\end{aligned}$$

As a result, rules in EAGLE give us the power to create specific temporal operators as well as to bind and modify data. This property turns out to be crucial for succinctly expressing executions of attack-safe systems. As we can see, each rule begins with a term that describes its type ( $\underline{\min}$ ,  $\underline{\max}$ ,  $\underline{\text{mon}}$ ); we will read more about it later.

Lastly, two assumptions must be made:

1. There is a finite sequence of events called  $\sigma = \alpha_1, \dots, \alpha_n$  that is a merge of the system registered logs organized by ascending time. The structure of an event record  $\alpha_i$  is the following:

$$\text{LoginLogoutEvent}\{\text{userId} : \underline{\text{string}}, \text{action} : \underline{\text{int}}, \text{time} : \underline{\text{double}}\}$$

An example of event could be:  $\{\text{userId} : \text{"Lori"}, \text{action} : \text{login}, \text{time} : 20\}$

2. For each attack, there is a formula  $F$  which specifies the absence of it.

To make the paper self-contained, we will present the syntax and semantics of EAGLE.

## 2.2 Syntax and Semantics

**Syntax.** A specification  $S$  is made up of an observer part  $O$  and a declaration part  $D$ .  $O$  is made up of zero or more monitor definitions  $M$ , which define what will be watched, while  $D$  is made up of zero or more rule definitions  $R$ . The names of rules and monitors are  $(N)$ .

$$\begin{aligned}
S &::= D O \\
D &::= R^* \\
O &::= M^* \\
R &::= \{\underline{\text{max}} \mid \underline{\text{min}}\} N(T_1 \ x_1, \dots, T_n \ x_n) = F \\
M &::= \underline{\text{mon}} N = F \\
T &::= \underline{\text{Form}} \mid \text{primitive type} \\
F &::= \text{expression} \mid \underline{\text{true}} \mid \underline{\text{false}} \mid \neg F \mid F_1 \wedge F_2 \mid F_1 \vee F_2 \mid \\
&F_1 \rightarrow F_2 \mid \bigcirc F \mid \odot F \mid F_1 \cdot F_2 \mid N(F_1, \dots, F_n) \mid x_i
\end{aligned}$$

The types  $T_i$  of parameters include primitive types like int, long, float, etc. or formulas of type Form. Definitions starting with keyword mon specifies the EAGLE formulas to be monitored and cannot have a recursive definition. We will see that these kind of rules will evolve as new events appear. A term indicating whether the interpretation is maximal (max) or minimal (min) comes before a rule specification  $R$ . While minimal rules establish *liveness properties* (something good eventually happens), maximal rules define *safety properties* (nothing bad ever happens); we shall see that the difference becomes important only when we are evaluating the at the boundaries of a trace. Finally, recursive definitions of rules are permitted as long as they are tightly guarded by a temporal operator.

**Semantics.** The semantics of the logic is defined in terms of the *satisfaction* relation  $\models$  which defines whether a finite execution trace  $\sigma$  satisfy the specification  $\varphi = D O$ . The  $i$ 'th state  $s_i$  of a trace  $\sigma$  is denoted by  $\sigma(i)$ . The term  $\sigma[i, j]$  denotes the sub-trace of  $\sigma$  from position  $i$  to position  $j$ , both positions included; if  $i \geq j$  then  $\sigma[i, j]$  denotes the empty trace.

Given a trace  $\sigma$  and a specification  $D O$ , satisfaction is defined as follows:

$$\sigma \models D O \text{ iff } \forall (\underline{\text{mon}} N = F) \in O. \sigma, 1 \models_D F$$

In other words, a trace satisfies a specification if it fulfills each monitored formula (mon) when monitoring from position 1, which is the initial state. The definition of the satisfaction relation  $\models_D \subseteq (\text{Trace} \times \mathbb{N}) \times \underline{\text{Form}}$ , for a set of rule definitions

$D$ , is the following:

$$\begin{aligned}
\sigma, i \models_D \text{expression} & \text{ iff } 1 \leq i \leq |\sigma| \text{ and } \text{evaluate}(\text{expression}, \sigma(i)) = \text{true} \\
\sigma, i \models_D \underline{\text{true}} & \\
\sigma, i \not\models_D \underline{\text{false}} & \\
\sigma, i \models_D \neg F & \text{ iff } \sigma, i \not\models_D F \\
\sigma, i \models_D F_1 \wedge F_2 & \text{ iff } \sigma, i \models_D F_1 \text{ and } \sigma, i \models_D F_2 \\
\sigma, i \models_D F_1 \vee F_2 & \text{ iff } \sigma, i \models_D F_1 \text{ or } \sigma, i \models_D F_2 \\
\sigma, i \models_D F_1 \rightarrow F_2 & \text{ iff } \sigma, i \models_D F_1 \text{ implies } \sigma, i \models_D F_2 \\
\sigma, i \models_D \bigcirc F & \text{ iff } i \leq |\sigma| \text{ and } \sigma, i+1 \models_D F \\
\sigma, i \models_D \odot F & \text{ iff } 1 \leq i \text{ and } \sigma, i-1 \models_D F \\
\sigma, i \models_D F_1 \cdot F_2 & \text{ iff } \exists j \text{ s.t. } i \leq j \leq |\sigma| + 1 \text{ and } \sigma^{[1, j-1]}, i \models_D F_1 \text{ and } \sigma^{[j, |\sigma|]}, 1 \models_D F_2 \\
\sigma, i \models_D N(\overline{F}, \overline{P}) & \text{ iff } \begin{cases} \text{if } 1 \leq i \leq |\sigma| \text{ then:} \\ \sigma, i \models_D F[\overline{f} \mapsto \overline{F}, \overline{p} \mapsto \text{evaluate}(\overline{P}, \sigma(i))] \\ \text{where } (N(\underline{\text{Form}} \overline{f}, \text{T } p) = F) \in D \\ \text{otherwise, if } i = 0 \text{ or } i = |\sigma| + 1 \text{ then:} \\ \text{rule } N \text{ is defined as } \underline{\text{max}} \text{ in } D \end{cases}
\end{aligned}$$

where  $\overline{F}$  and  $\overline{P}$  represent tuples of type  $\overline{\text{Form}}$  and  $\overline{T}$  respectively

An **important** note is that, given a trace  $\sigma = \alpha_1, \dots, \alpha_n$  the index  $i$  of the trace can become 0 (before the first state) and  $n + 1$ , thus going beyond the limits; both of these situations result in rule applications evaluating to either true if maximal or false if minimal, without first taking the rules body into account. Let us explain why using an example. Given the definition

$$\underline{\text{min}} \text{ Property}(\underline{\text{Form}} F) = F \vee \bigcirc \text{Property}(F)$$

and the sequence  $\sigma = \alpha_1, \dots, \alpha_n$ , Property will evaluate to *true*

$$\text{evaluate}(\dots \text{evaluate}(\text{evaluate}(\text{Property}, \alpha_1), \alpha_2), \dots, \alpha_n) = \text{true}$$

if and only if the rule is true at some given event  $\alpha_i$ . Once the sequence has been completely analyzed, we will obtain a big disjunction (in case of a min) or a big conjunction (in case of a max) and the final value will close the evaluation. Continuing the example, the extended formula will be:

$$F \vee \bigcirc (F \vee \bigcirc (\dots F \vee \bigcirc \mathbf{Property}(\mathbf{F})))$$

and the bold **Property(F)** will be evaluated to false, since it is the last of a big disjunction.

## 2.3 Evaluation algorithm

As already mentioned, monitored formulas (mon) evolve at every event, to store relevant informations about past instants. The evolved formula's value is determined at the conclusion of the event series; if it is true, the formula is satisfied

by the event sequence; otherwise, it is violated. Formally, a formula  $F' = \text{evaluate}(F, \alpha_i)$  is created when a formula  $F$  is evaluated at an event  $\alpha_i = \sigma(i)$  and has the condition that  $\sigma, i \models F$  if and only if  $\sigma, i+1 \models F'$ . With  $F$  being the evolved formula, we compute the boolean function  $\text{value}(F)$  at the conclusion of the trace. The definition of the function  $\text{evaluate} : \underline{\text{Form}} \times \underline{\text{State}} \rightarrow \underline{\text{Form}}$  uses an auxiliary function  $\text{update} : \underline{\text{Form}} \times \underline{\text{State}} \rightarrow \underline{\text{Form}}$  to pre-evaluate a formula if it is guarded by the previous operator  $\odot$ . At the end (or at the beginning) of a trace, the function  $\text{value} : \underline{\text{Form}} \rightarrow \{\underline{\text{true}}, \underline{\text{false}}\}$ , when applied to  $F$ , returns true if and only if  $\sigma, |\sigma| + 1 \models F$  (or  $\sigma, 0 \models F$ ), and returns false otherwise.

**Calculus.** According to [3], the  $\text{evaluate}$ ,  $\text{update}$  and  $\text{value}$  functions are defined a priori for all operators except for the rule application, for which get generated based on the definition of rules in the specification. The definitions of these functions on the different primitive operators are the following ( $op$  can be  $\vee, \wedge, \rightarrow$ ):

$$\begin{aligned}
& \text{evaluate}(\underline{\text{true}}, \alpha_i) = \underline{\text{true}} \\
& \text{evaluate}(\underline{\text{false}}, \alpha_i) = \underline{\text{false}} \\
& \text{evaluate}(\text{exp}, \alpha_i) = \text{value of exp in } \alpha_i \\
& \text{evaluate}(F_1 \text{ op } F_2, \alpha_i) = \text{evaluate}(F_1, \alpha_i) \text{ op } \text{evaluate}(F_2, \alpha_i) \\
& \text{evaluate}(\neg F, \alpha_i) = \neg \text{evaluate}(F, \alpha_i) \\
& \text{evaluate}(\odot F, \alpha_i) = \text{update}(F, \alpha_i) \\
& \text{evaluate}(F_1 \cdot F_2, \alpha_i) = \begin{cases} \text{evaluate}(F_1, \alpha_i) \cdot F_2 & \text{if } \text{value}(F_1) = \underline{\text{false}} \\ \text{evaluate}(F_1, \alpha_i) \cdot F_2 \vee \text{evaluate}(F_2, \alpha_i) & \text{else} \end{cases}
\end{aligned}$$

$$\begin{aligned}
& \text{value}(\underline{\text{true}}) = \underline{\text{true}} \\
& \text{value}(\underline{\text{false}}) = \underline{\text{false}} \\
& \text{value}(\text{exp}) = \underline{\text{false}} \\
& \text{value}(F_1 \text{ op } F_2) = \text{value}(F_1) \text{ op } \text{value}(F_2) \\
& \text{value}(\neg F) = \neg \text{value}(F) \\
& \text{value}(\odot F) = \begin{cases} F & \text{if at the beginning of a trace} \\ \underline{\text{false}} & \text{if at the end of a trace} \end{cases} \\
& \text{value}(F_1 \cdot F_2, \alpha_i) = \text{value}(F_1) \wedge \text{value}(F_2)
\end{aligned}$$

$$\begin{aligned}
& \text{update}(\underline{\text{true}}, \alpha_i) = \underline{\text{true}} \\
& \text{update}(\underline{\text{false}}, \alpha_i) = \underline{\text{false}} \\
& \text{update}(\text{exp}, \alpha_i) = \text{exp} \\
& \text{update}(F_1 \text{ op } F_2, \alpha_i) = \text{update}(F_1, \alpha_i) \text{ op } \text{update}(F_2, \alpha_i) \\
& \text{update}(\neg F, \alpha_i) = \neg \text{update}(F, \alpha_i) \\
& \text{update}(\odot F, \alpha_i) = \text{update}(F, \alpha_i)
\end{aligned}$$

$$\text{update}(F_1 \cdot F_2, \alpha_i) = \text{update}(F_1, \alpha_i) \cdot F_2$$

Let us introduce an example to better understand the algorithm. Given the following rules

$$\begin{aligned} \underline{\max} \text{ Always}(\text{Form } F) &= F \wedge \bigcirc \text{Always}(F) \\ \underline{\min} \text{ EvTimedLogout}(\text{string } k, \text{double } t, \text{double } \delta) &= (\text{time} - t \leq \delta) \\ &\quad \wedge ((\text{action} = \text{logout} \wedge \text{userId} = k) \vee \bigcirc \text{EvTimedLogout}(k, t, \delta)) \\ \underline{\text{mon}} M_3 &= \text{Always}((\text{action} = \text{login}) \rightarrow \text{EvTimedLogout}(\text{userId}, \text{time}, 100)) \end{aligned}$$

and a trace  $\sigma = \alpha_1, \alpha_2$ , where  $\alpha_1 = \{\text{userId} : \text{"Lori"}, \text{action} : \text{login}, \text{time} : 17.0\}$  and  $\alpha_2 = \{\text{userId} : \text{"Lori"}, \text{action} : \text{logout}, \text{time} : 150.0\}$ , the initial formula  $M_3$  gets modified. At time 1, when  $\alpha_1$  is analyzed,  $M_3$  changes as follows:

$$\begin{aligned} \text{evaluate}(M_3, \alpha_1) &= \text{EvTimedLogout}(\text{"Lori"}, 17.0, 100) \\ &\quad \wedge \text{Always}((\text{action} = \text{login}) \rightarrow \text{EvTimedLogout}(\text{userId}, \text{time}, 100)) \end{aligned}$$

At the second event the predicate  $(\text{time} - t \leq \delta)$  gets instantiated to  $(150.0 - 17.0 \leq 100)$  which is false; thus the whole formula becomes false, indicating that the two-event trace violates the property  $M_3$ .

It's essential to emphasize that the logic EAGLE operates with a *finite* trace semantics; the monitoring algorithm, as written before, can check whether the formula is satisfied or not only at the conclusion of a trace. However, in intrusion detection scenarios, the notion of the "end of a trace" makes no sense since event sequences can theoretically be infinite. In such cases, the aim is to trigger an alarm as soon as a property is violated and this is done by continuously checking the formula's satisfaction status after each event. In practice, once we have the specification of the attack pattern in terms of  $\varphi$ , we can express the system's safe behavior as  $\Box(\neg\varphi)$ . Note that if a sequence of events corresponding to the attack is discovered, this formula becomes false.

## 2.4 Relationship to LTL+P

According to [3, 1], we can define both **Future Time LTL** and **Past Time LTL** operators. The future modalities are written in section 2.1, therefore let us list the past ones:

$$\begin{aligned} \underline{\min} \text{ Previous}(\text{Form } F) &= \odot F \\ \underline{\max} \text{ AlwaysInPast}(\text{Form } F) &= F \wedge \odot \text{AlwaysInPast}(F) \\ \underline{\min} \text{ EventuallyInPast}(\text{Form } F) &= F \vee \odot \text{EventuallyInPast}(F) \\ \underline{\min} \text{ Since}(\text{Form } F_1, \text{Form } F_2) &= F_2 \vee (F_1 \wedge \odot \text{Since}(F_1, F_2)) \end{aligned}$$

By combining the definitions for the future and past time LTLs defined above, we obtain a temporal logic over the future, present and past, in which one can freely intermix the future and past time modalities.

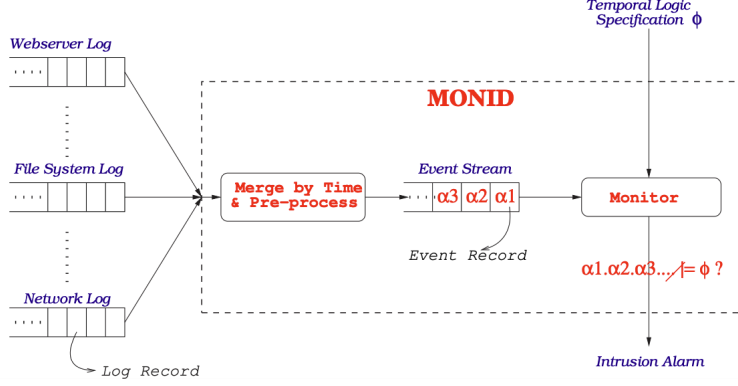


Figure 1: MONID Architecture.

### 3 MONID: Structure analysis

MONID is a prototype which can detect intrusions either online or offline. System-level event data is delivered to MONID server either directly from relevant log files (offline) or online via application code that has been properly instrumented. In the online mode, MONID runs as a server that receives streams of events from various sources and to generate these events, the different logging modules are instrumented so that they filter and send relevant events to the server. To create a single event trace, the server combines events from numerous sources according to timestamps and preprocesses them into an abstract intermediate form. Note that sometimes it is necessary to watch events from multiple sources in order to detect certain attacks; then, MONID monitor checks this event trail against a predetermined criteria, and if the specification is broken, an intrusion alarm is raised. While, in the offline mode, MONID reads various log files and sends an event corresponding each log entry to the server. The server then processes the event stream as before to detect intrusion.

The entire framework is represented in Figure 1. In paper [8], the implementation was written in Java and available as a library; there were two available methods used. The first one is *parse* which takes a file containing a list of EA-GLE specifications and monitor formulas and compiles them into a data structure used to represent the *monitor* represented in Figure 1. The client software iteratively invokes the method *evaluate* for each event after compilation; this call internally adjusts the monitor formulas in accordance with Subsection 2.3. An error notice is produced or a pre-specified mechanism is triggered to take a corrective action if the monitored formulas become false at any stage.

Some evaluations were also have been done in [8], using the DARPA Intrusion Detection Evaluation dataset [7]. The experiment was done using offline mode, against some of the attacks presented in Section 4. Although the used hardware is now outdated and obsolete, the overhead was rather modest, thus this method should still be helpful in theory today.



## 4 Example of attacks

## 5 Conclusions

## References

- [1] Howard Barringer et al. “EAGLE can do Efficient LTL Monitoring”. In: *Fossacs Ortacas*. 2003.
- [2] Howard Barringer et al. “Program monitoring with LTL in EAGLE”. In: *18th International Parallel and Distributed Processing Symposium, 2004. Proceedings*. IEEE. 2004, p. 264.
- [3] Howard Barringer et al. “Rule-based runtime verification”. In: *Verification, Model Checking, and Abstract Interpretation: 5th International Conference, VMCAI 2004 Venice, Italy, January 11-13, 2004 Proceedings 5*. Springer. 2004, pp. 44–57.
- [4] Wei Gao and Thomas H Morris. “On cyber attacks and signature based intrusion detection for modbus based industrial control systems”. In: *Journal of Digital Forensics, Security and Law* 9.1 (2014), p. 3.
- [5] Akash Garg and Prachi Maheshwari. “A hybrid intrusion detection system: A review”. In: *2016 10th International Conference on Intelligent Systems and Control (ISCO)*. IEEE. 2016, pp. 1–5.
- [6] VVRPV Jyothsna, Rama Prasad, and K Munivara Prasad. “A review of anomaly based intrusion detection systems”. In: *International Journal of Computer Applications* 28.7 (2011), pp. 26–35.
- [7] John McHugh. “Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory”. In: *ACM Transactions on Information and System Security (TISSEC)* 3.4 (2000), pp. 262–294.
- [8] Prasad Naldurg, Koushik Sen, and Prasanna Thati. “A temporal logic based framework for intrusion detection”. In: *Formal Techniques for Networked and Distributed Systems-FORTE 2004: 24th IFIP WG 6.1 International Conference, Madrid Spain, September 27-30, 2004. Proceedings 24*. Springer. 2004, pp. 359–376.