Recognizing Unexplained Behavior in Network Traffic*

Massimiliano Albanese, Robert F. Erbacher, Sushil Jajodia, Cristian Molinaro, Fabio Persia, Antonio Picariello, Giancarlo Sperlì, and V. S. Subrahmanian

Abstract Intrusion detection and alert correlation are valuable and complementary techniques for identifying security threats in complex networks. Intrusion detection systems monitor network traffic for suspicious behavior, and trigger security alerts. Alert correlation methods can aggregate such alerts into multi-step attacks scenarios. However, both methods rely on models encoding a priori knowledge of either normal or malicious behavior. As a result, these methods are incapable of quantifying how well the underlying models explain what is observed on the network. To overcome this limitation, we present a framework for evaluating the probability that a sequence of events is not explained by a given a set of models. We leverage important properties of this framework to estimate such probabilities efficiently, and design fast algorithms for identifying sequences of events that are unexplained with a probability above a given threshold. Our framework can operate both at the intrusion detection level and at the alert correlation level. Experiments on a prototype implementation of the framework show that our approach scales well and provides accurate results.

Massimiliano Albanese and Sushil Jajodia
George Mason University, Fairfax, USA e-mail: {malbanes,jajodia}@gmu.edu

Robert F. Erbacher
U.S. Army Research Laboratory, USA e-mail: robert.f.erbacher.civ@mail.mil

Cristian Molinaro
University of Calabria, Rende, Italy e-mail: cmolinaro@deis.unical.it

Fabio Persia, Antonio Picariello, and Giancarlo Sperlì
University of Naples Federico II, Naples, Italy e-mail: {fabio.persia,picus,g.sperli}@unina.it

V. S. Subrahmanian
University of Maryland, College Park, USA e-mail: vs@umiacs.umd.edu

* The work presented in this paper is supported in part by the Army Research Office under MURI award number W911NF-09-1-05250525, and by the Office of Naval Research under award number N00014-12-1-0461. Part of the work was performed while Sushil Jajodia was a Visiting Researcher at the US Army Research Laboratory.

1 Introduction

Intrusion detection and alert correlation techniques provide valuable and complementary tools for identifying and monitoring security threats in complex network infrastructures. Intrusion detection systems (IDS) can monitor network traffic for suspicious behavior and trigger security alerts accordingly [8, 6, 7]. Alert correlation methods can aggregate such alerts into multi-step attack scenarios [15, 10, 5, 9, 1].

Intrusion detection has been studied for about thirty years, since it was first identified in the Anderson report [4], and it is based on the assumption that an intruder's behavior will be noticeably different from that of a legitimate user and that many unauthorized actions are detectable [8].

Intrusion detection techniques can be broadly classified into *signature-based* [7] and *profile-based* (or *anomaly-based*) [6] methods. There are advantages and disadvantages to each method. The trend today is to use the two methods together to provide the maximum defense for the network infrastructure. A signature generally refers to a set of conditions that characterize the direct manifestation of intrusion activities in terms of packet headers and payload content. Historically, signature-based methods have been most common when looking for suspicious or malicious activity on the network. These methods rely on a database of attack signatures and when one or more of these signatures match what is observed in live traffic, an alarm is triggered and the event is logged for further investigation. The effectiveness of signature-based intrusion detection is highly dependent on its signature database. If a signature is not in the database, the IDS will not recognize the attack.

Anomaly-based intrusion detection triggers an alarm when some type of unusual behavior occurs on the network. This would include any event that is considered to be abnormal by a predefined standard. Anything that deviates from this baseline of *normal* behavior will be flagged and logged as *anomalous* or *suspicious*. For instance, HTTP traffic on a non-standard port, say port 63, would be flagged as suspicious. Normal behavior can be programmed into the system based on offline learning and research, or the system can learn the normal behavior online while processing the network traffic.

In complex networks, most intrusions are not isolated but represent different stages of specific attack sequences, with the early stages preparing for the later ones. In other words, complex attacks consist of multiple steps, each triggering specific security alerts. This fundamental observation, along with the potentially large number of alerts deriving from the widespread deployment of IDS sensors, has prompted significant research in automatic alert correlation techniques. The goal of correlation is to find causal relationships between alerts in order to reconstruct attack scenarios from isolated alerts [14]. Although it may not significantly reduce the number of alerts, the main role of correlation is to provide a higher level view of the actual attacks. Existing approaches to alert correlation can be divided into the following categories based on the criterion used for relating alerts: scenario-based correlation [1, 5], rule-based correlation [9], statistical correlation [12, 13], and temporal correlation [11].

From a conceptual point of view, both intrusion detection systems and alert correlation methods aggregate fine grained information into higher level views of the attack, although they operate at different levels of abstraction, as shown in Figure 1. Moreover, both rely on models encoding a priori knowledge of either normal or malicious behavior, and cannot appropriately deal with events that are not *explained* by the underlying models. In practice, all these methods are incapable of quantifying how well available models explain a sequence of events observed in data streams (data packets and alerts respectively) feeding the two classes of tools.

To address this limitation and offer novel analytic capabilities, we present a framework for evaluating the probability that a sequence of events – either at the network traffic level or at the alert level – is unexplained, given a set of models of previous learned behaviors. Our approach is an application of the framework proposed in [3] to the cyber security setting. We adapt algorithms from [3] so as to efficiently *estimate* the probability that a sequence is unexplained (rather than computing the exact probability as done in [3]). The computation of approximate probabilities is done by leveraging the mathematical properties studied in Section 3. Our framework can operate both at the intrusion detection level and at the alert correlation level, but it is not intended to replace existing tools. In fact, our framework builds on top of these tools, and analyzes their output in order to identify what is not "sufficiently" explained by the underlying models. Experiments on a prototype implementation of the framework show that our approach scales well and provides accurate results.

The rest of the paper is organized as follows. Section 2 presents the proposed probabilistic model, whereas Section 3 discusses the properties that can be leveraged to compute approximate probabilities efficiently. Efficient algorithms to recognize unexplained behaviors are presented in Section 4. An experimental evaluation of our framework is reported in Section 5, and concluding remarks are given in Section 6.

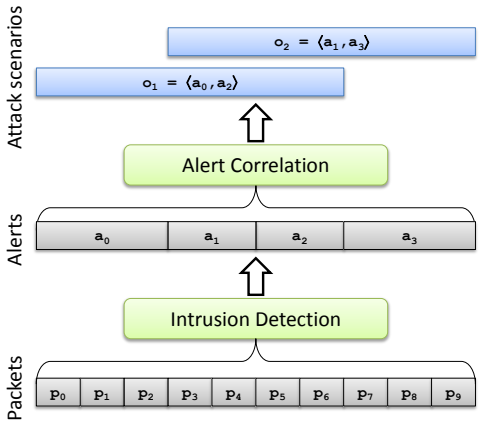


Fig. 1 Conceptual diagram of the relationship between alert correlation and intrusion detection

2 Behavior Model

In this section, we present a framework for evaluating the probability that a sequence of events is unexplained, given a set of models. As already mentioned above, this framework can operate both at alert correlation level and at intrusion detection level. Note that the model described in this section has been adapted from previous work on activity detection for video surveillance applications [3]. The novel contribution of this paper starts in Section III, where we propose efficient techniques to compute an approximate probability that a sequence of events is unexplained.

2.1 Preliminaries

We assume the existence of a set $\mathscr{E}$ of observable *events*[2], and a set $\mathscr{A}$ of *models*[3] representing known behavior (either normal or malicious) in terms of observable events. When an event is observed, an *observation* is generated. An observation is a pair $a = (e,ts)$, where $e \in \mathscr{E}$ is an observable event, and ts is a timestamp recording the time at which an instance of e was observed. An *observation stream* (or *observation sequence*) S is a finite sequence of observations.

Example 1. Consider the Snort rule `alert any any -> any any (flags:SF,12; msg:``Possible SYN FIN scan'';)`[4]. This rule detects when a packet has the SYN and FIN flags set at the same time – indicating a possible SYN FIN scan attempt – and generates an alert (observation) $a = (e,ts)$, where the observable event e is the fact that the SYN and FIN flags are set at the same time, and ts is the time at which the packet was observed.

Throughout the paper, we use the following terminology and notation for sequences. Let $S_1 = \langle a_1,\ldots,a_n\rangle$ and $S_2 = \langle b_1,\ldots,b_m\rangle$ be two sequences. We say that S_2 is a *subsequence* of S_1 iff there exist $1 \leq j_1 < j_2 < \cdots < j_m \leq n$ s.t. $b_i = a_{j_i}$ for $1 \leq i \leq m$. If $j_{i+1} = j_i + 1$ for $1 \leq i < m$, then S_2 is a *contiguous* subsequence of S_1. We write $S_1 \cap S_2 \neq \emptyset$ iff S_1 and S_2 have a common element and write $e \in S_1$ iff e is an element appearing in S_1. The *concatenation* of S_1 and S_2, i.e., the sequence $\langle a_1,\ldots,a_n,b_1,\ldots,b_m\rangle$, is denoted by $S_1 \cdot S_2$. Finally, we use $|S_1|$ to denote the length of S_1, that is, the number of elements in S_1.

Given an observation stream S, and a behavior model $A \in \mathscr{A}$, an *occurrence* o of A in S is a subsequence $\langle (e_1,ts_1),\ldots,(e_m,ts_m)\rangle$ of S such that the sequence of events $\langle e_1,\ldots,e_m\rangle$ represents a possible way of exhibiting behavior A (e.g., a specific path in the attack graph from initial to target conditions). The relative probability[5] $p^*(o)$

[2] At the intrusion detection level, observable events may simply be observable packet features. At the alert correlation level, observable events are alerts generated by the underlying intrusion detection system.

[3] At the intrusion detection level, $\mathscr{A}$ is a set of IDS rules. At the alert correlation level, $\mathscr{A}$ is a set of attack models, such as attack graphs.

[4] `http://www.snort.org/`

[5] Probabilities of occurrences must be normalized in order to enable comparison of occurrences of different behavior models.

of the occurrence o is the probability that the sequence of events $\langle e_1, \ldots, e_m \rangle$ is in fact an instance of the corresponding behavior. The problem of computing the probability of an occurrence is beyond the scope of our paper. However, several researchers have addressed this problem. For instance, a probabilistic extension of attack graphs is proposed in [2], along with an algorithm for identifying attack occurrences efficiently. Therefore, we assume that all the occurrences o of a behavior A and their probabilities can be readily computed.

We use $\mathscr{O}(S, \mathscr{A})$ to denote the set of all occurrences in S of behaviors in $\mathscr{A}$. When S and $\mathscr{A}$ are clear from the context, we write $\mathscr{O}$ instead of $\mathscr{O}(S, \mathscr{A})$.

2.2 Probabilistic Unexplained Behavior Model

We now define the probability that an observation sequence is unexplained, given a set $\mathscr{A}$ of known behaviors. We start by noting that the probabilistic nature of *occurrences* implicitly involves conflicts. For instance, consider the two occurrences o_1 and o_2 in Figure 2. In this case, there is an implicit conflict because a_2 belongs to both occurrences, but in fact, a_2 can only belong to one occurrence, i.e., though o_1 and o_2 may both have a non-zero probability of occurrence, the probability that they coexist is 0[6]. Formally, we say two occurrences o_i, o_j *conflict*, denoted $o_i \nsim o_j$, iff $o_i \cap o_j \neq \emptyset$. We now use this notion to define possible worlds.

Definition 1 (Possible world). A *possible world* for an observation sequence S w.r.t. a set of behavior models $\mathscr{A}$ is a subset w of $\mathscr{O}(S, \mathscr{A})$ s.t. $\nexists o_i, o_j \in w, o_i \nsim o_j$.

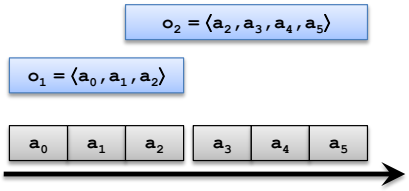


Fig. 2 Example of conflicting occurrences

Thus, a possible world is a set of occurrences which do not conflict with one another, i.e., an observation cannot be identified as part of two distinct occurrences in the same world. We use $\mathscr{W}(S, \mathscr{A})$ to denote the set of all possible worlds for an observation sequence S w.r.t. a set of behavior models $\mathscr{A}$; when S and $\mathscr{A}$ are clear from context, we write $\mathscr{W}$.

[6] This assumption makes modeling simpler, but it can be removed or modified in situations where certain atomic events are shared among multiple attack patterns.

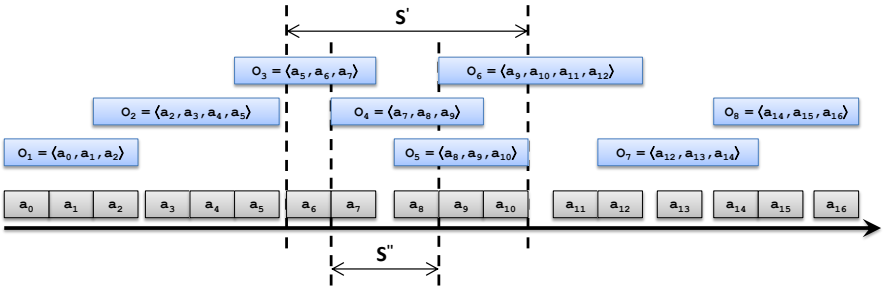


Fig. 3 Example of observation sequence and occurrences

Example 2. Consider the observation sequence and the two conflicting occurrences o_1, o_2 in Figure 2. There are 3 possible worlds: $w_0 = \emptyset$, $w_1 = \{o_1\}$, and $w_2 = \{o_2\}$. Note that $\{o_1, o_2\}$ is not a world as $o_1 \nsim o_2$. Each world represents a way of explaining what is observed. The first world corresponds to the case where nothing is explained, the second and third worlds correspond to the scenarios where we use one of the two possible occurrences to explain the observed events.

Note that any subset of $\mathscr{O}$ not containing conflicting occurrences is a legitimate possible world – possible worlds are not required to be maximal w.r.t. set inclusion $\subseteq$. In the above example, the empty set is a possible world even though there are two other possible worlds $w_1 = \{o_1\}$ and $w_2 = \{o_2\}$ which are supersets of it. The reason is that o_1 and o_2 are uncertain, so the scenario where neither o_1 nor o_2 occurs is a legitimate one. We further illustrate this point below.

Example 3. Consider an observation sequence where a single occurrence o has been identified with $p^*(o) = 0.6$. In this case, it is natural to say that there are two possible worlds $w_0 = \emptyset$ and $w_1 = \{o\}$ and expect the probabilities of w_0 and w_1 to be 0.4 and 0.6, respectively. By restricting ourselves to maximal possible worlds only, we would have only one possible world, w_1, whose probability is 1, which is wrong. Nevertheless, if $p^*(o) = 1$, w_1 is the only possible scenario. This can be achieved by assigning 0 and 1 to the probabilities of w_0 and w_1, respectively.

Thus, occurrences determine a set of possible worlds (intuitively, different ways of explaining the observation stream). We wish to find a probability distribution over all possible worlds that (i) is consistent with the relative probabilities of the occurrences, and (ii) takes conflicts into account. We assume the user specifies a function $\omega : \mathscr{A} \to \mathbb{R}^+$ which assigns a weight to each behavior and prioritizes the importance of the behavior[7]. The weight of an occurrence o of behavior A is the weight of A.

We use $C(o)$ to denote the set of occurrences conflicting with o, i.e., $C(o) = \{o' \mid o' \in \mathscr{O} \wedge o' \nsim o\}$. Note that $o \in C(o)$; furthermore, $C(o) = \{o\}$ when o does not

[7] For instance, highly threatening behaviors may be assigned a high weight.

conflict with any other occurrence. Suppose p_i denotes the (unknown) probability of world w_i. As we know the probability of occurrences, and as each occurrence occurs in certain worlds, we can induce a set of linear constraints that can be used to learn the values of the p_i's.

Definition 2. Let S be an observation sequence, $\mathscr{A}$ a set of behavior models, and $\mathscr{O}$ the set of occurrences identified in S w.r.t. $\mathscr{A}$. We define the linear constraints $\mathsf{LC}(S,\mathscr{A})$ as follows:

$$\begin{cases} p_i \geq 0, \quad \forall w_i \in \mathscr{W} \\ \sum\limits_{w_i \in \mathscr{W}} p_i = 1 \\ \sum\limits_{w_i \in \mathscr{W}\, s.t.\, o \in w_i} p_i = p^*(o) \cdot \dfrac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)}, \forall o \in \mathscr{O} \end{cases}$$

The first two types of constraints enforce a probability distribution over the set of possible worlds. The last type of constraint ensures that the probability of occurrence o – which is the sum of the probabilities of the worlds containing o – is equal to its relative probability $p^*(o)$ weighted by $\frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)}$, the latter being the weight of o divided by the sum of the weights of the occurrences conflicting with o. Note that: (i) the value on the right-hand side of the last type of constraint decreases as the amount of conflict increases, (ii) if an occurrence o is not conflicting with any other occurrence, then its probability $\sum_{w_i \in \mathscr{W}\, s.t.\, o \in w_i} p_i$ is equal to $p^*(o)$, i.e. the relative probability returned by chosen tool for identifying behavior occurrences in observation streams.

Example 4. Consider the observation sequence and occurrences of Figure 3, and assume that such occurrences have been identified with the relative probabilities shown in the second column of Table 1. The table also shows the weights assigned to the occurrences and the value of $p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)}$. The 8 occurrences determine 49 possible worlds[8]. The set of linear constraints of Definition 2 for this case is shown in Figure 4.

In the rest of the paper, we assume that $\mathsf{LC}(S,\mathscr{A})$ is solvable. We give two semantics for a subsequence S' of an observation sequence S to be unexplained in a world $w \in \mathscr{W}$:

1. S' is *totally unexplained* in w, denoted $w \not\models_T S'$, iff $\forall a_i \in S', \nexists o \in w, a_i \in o$;
2. S' is *partially unexplained* in w, denoted $w \not\models_P S'$, iff $\exists a_i \in S', \nexists o \in w, a_i \in o$.

Intuitively, S' is totally (resp. partially) unexplained in w iff w does not explain every (resp. at least one) observation in S'.

When we have a probability distribution over the set of possible worlds (i.e., a solution of $\mathsf{LC}(S,\mathscr{A})$), the probability that a sequence S' is totally (resp. partially)

[8] We do not list all the worlds for reason of space.

o	$p^*(o)$	$\omega(o)$	$p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)}$
o_1	0.90	3	0.45
o_2	0.80	3	0.30
o_3	0.72	2	0.16
o_4	0.65	4	0.20
o_5	0.77	4	0.28
o_6	0.50	3	0.10
o_7	0.60	4	0.24
o_8	0.70	3	0.30

Table 1 Probabilities and weights of occurrences

$$
\begin{cases}
p_0 \geq 0 \\
p_1 \geq 0 \\
\cdots \\
p_{48} \geq 0 \\
p_0 + p_1 + \cdots + p_{48} = 1 \\
p_1 + p_9 + p_{10} + p_{11} + p_{12} + p_{13} + p_{14} + p_{28} + p_{29} + p_{30} + p_{31} + p_{32} + p_{33} + p_{34} + p_{35} + p_{36} + p_{45} + p_{46} + p_{47} + p_{48} = 0.45 \\
p_2 + p_{15} + p_{16} + p_{17} + p_{18} + p_{19} + p_{38} + p_{39} + p_{40} + p_{41} + p_{42} = 0.3 \\
p_3 + p_9 + p_{20} + p_{21} + p_{22} + p_{23} + p_{29} + p_{30} + p_{31} + p_{32} + p_{43} + p_{44} + p_{45} + p_{46} + p_{47} + p_{48} = 0.16 \\
p_4 + p_{10} + p_{15} + p_{24} + p_{25} + p_{33} + p_{34} + p_{38} + p_{39} = 0.2 \\
p_5 + p_{11} + p_{16} + p_{20} + p_{26} + p_{27} + p_{29} + p_{35} + p_{36} + p_{40} + p_{41} + p_{43} + p_{44} + p_{46} + p_{47} = 0.28 \\
p_6 + p_{12} + p_{17} + p_{21} + p_{28} + p_{30} + p_{37} + p_{42} + p_{45} + p_{48} = 0.1 \\
p_7 + p_{13} + p_{18} + p_{22} + p_{24} + p_{26} + p_{31} + p_{33} + p_{35} + p_{38} + p_{40} + p_{43} + p_{46} = 0.24 \\
p_8 + p_{14} + p_{19} + p_{23} + p_{25} + p_{27} + p_{28} + p_{32} + p_{34} + p_{36} + p_{37} + p_{39} + p_{41} + p_{42} + p_{44} + p_{45} + p_{47} + p_{48} = 0.3
\end{cases}
$$

Fig. 4 Linear constraints for the occurrences of Figure 3

unexplained can be naturally defined as the sum of the probabilities of the worlds w_i s.t. $S^{'}$ is totally (resp. partially) unexplained in w_i. This is formally defined as follows.

Definition 3. Let $\mathscr{A}$ be a set of behavior models, S an observation sequence, and $S^{'}$ a subsequence of S. Suppose we have a probability distribution ϕ over $\mathscr{W}$ obtained by solving $\mathsf{LC}(S, \mathscr{A})$. The probability that $S^{'}$ is totally unexplained in S w.r.t. $\mathscr{A}$ and ϕ is

$$\mathscr{P}_T(S^{'}, \mathscr{A}, \phi) = \sum_{w_i \in \mathscr{W} \ s.t. w_i \not\models_T S^{'}} \phi(w_i)$$

Similarly, the probability that $S^{'}$ is partially unexplained in S w.r.t. $\mathscr{A}$ and ϕ is

$$\mathscr{P}_P(S^{'}, \mathscr{A}, \phi) = \sum_{w_i \in \mathscr{W} \ s.t. w_i \not\models_P S^{'}} \phi(w_i)$$

The previous definition gives the probability that a sequence $S^{'}$ is totally (resp. partially) unexplained for a given solution of $\mathsf{LC}(S, \mathscr{A})$. However, in general $\mathsf{LC}(S, \mathscr{A})$ can admit multiple solutions, each yielding a probability that a sequence is totally or partially unexplained. We define the probability interval that a sequence $S^{'}$ is totally (resp. partially) unexplained by minimizing and maximizing the probability that $S^{'}$ is totally (resp. partially) unexplained subject to the linear constraints of Definition 2.

Definition 4. Let $\mathscr{A}$ be a set of behavior models, S an observation sequence, and $S^{'}$ a subsequence of S. The probability interval that $S^{'}$ is totally unexplained in S w.r.t. $\mathscr{A}$ is $\mathscr{I}_T(S^{'},\mathscr{A}) = [l,u]$, where:

$$l = \textbf{minimize} \sum_{w_i \in \mathscr{W}\ s.t. w_i \not\models_T S'} p_i \quad \textbf{subject to } \mathsf{LC}(S,\mathscr{A})$$

$$u = \textbf{maximize} \sum_{w_i \in \mathscr{W}\ s.t. w_i \not\models_T S'} p_i \quad \textbf{subject to } \mathsf{LC}(S,\mathscr{A})$$

Likewise, the probability interval that $S^{'}$ is partially unexplained in S w.r.t. $\mathscr{A}$ is $\mathscr{I}_P(S^{'},\mathscr{A}) = [l,u]$, where:

$$l = \textbf{minimize} \sum_{w_i \in \mathscr{W}\ s.t. w_i \not\models_P S'} p_i \quad \textbf{subject to } \mathsf{LC}(S,\mathscr{A})$$

$$u = \textbf{maximize} \sum_{w_i \in \mathscr{W}\ s.t. w_i \not\models_P S'} p_i \quad \textbf{subject to } \mathsf{LC}(S,\mathscr{A})$$

Thus, the probability $\mathscr{P}_T(S^{'},\mathscr{A},\phi)$ (resp. $\mathscr{P}_P(S^{'},\mathscr{A},\phi)$) that a subsequence $S^{'}$ of S is totally (resp. partially) unexplained w.r.t. a solution ϕ of $\mathsf{LC}(S,\mathscr{A})$ is the sum of the probabilities of the worlds in which $S^{'}$ is totally (resp. partially) unexplained. As $\mathsf{LC}(S,\mathscr{A})$ may have multiple solutions, we find the tightest interval $[l,u]$ (resp. $[l',u']$) containing this probability for any solution. Different criteria can be used to choose a point probability value from an interval $[l,u]$, e.g., the minimum (l), the maximum (u), or the average (i.e., $(l+u)/2$).

In the rest of the paper we assume that one of the above criteria has been chosen, and we use $\mathscr{P}_T(S^{'},\mathscr{A})$ (resp. $\mathscr{P}_P(S^{'},\mathscr{A})$) to denote the probability that $S^{'}$ is totally (resp. partially) unexplained; when $\mathscr{A}$ is clear from context, we write $\mathscr{P}_T(S^{'})$ (resp. $\mathscr{P}_P(S^{'})$).

Example 5. Consider the observation sequence and occurrences of Figure 3. The probability $\mathscr{P}_T(S^{'})$ that the sequence $S^{'} = \langle a_6,a_7,a_8,a_9,a_{10}\rangle$ is totally unexplained is obtained by minimizing and maximizing the objective function $\sum_{w_i \in \mathscr{W}\ s.t. w_i \not\models_T S'} p_i = p_0+p_1+p_2+p_7+p_8+p_{13}+p_{14}+p_{18}+p_{19}$ subject to the constraints of Figure 4, which gives $\mathscr{I}_T(S^{'}) = [0.26,0.42]$. The probability $\mathscr{P}_P(S^{''})$ that the sequence $S^{''} = \langle a_7,a_8\rangle$ is partially unexplained is obtained by minimizing and maximizing the corresponding objective function[9] which gives $\mathscr{I}_P(S^{'}) = [0.64,0.8]$.

Proposition 1. *Consider two subsequences S_1 and S_2 of an observation sequence S. If S_1 is a subsequence of S_2, then $\mathscr{P}_T(S_1) \geq \mathscr{P}_T(S_2)$ and $\mathscr{P}_P(S_1) \leq \mathscr{P}_P(S_2)$.*

We now define totally and partially unexplained behaviors.

[9] This objective function is the sum of 34 variables and is not shown for reasons of space.

Definition 5 (Unexplained behavior). Let S be an observation sequence, $\tau \in [0,1]$ a probability threshold, and $L \in \mathbb{N}^+$ a length threshold. Then,

- a *totally unexplained behavior* is a subsequence $S^{'}$ of S s.t. (i) $\mathscr{P}_T(S^{'}) \geq \tau$, (ii) $|S^{'}| \geq L$, and (iii) $S^{'}$ is maximal, i.e., there does not exist a subsequence $S^{''} \neq S^{'}$ of S s.t. $S^{'}$ is a subsequence of $S^{''}$, $\mathscr{P}_T(S^{''}) \geq \tau$, and $|S^{''}| \geq L$.
- a *partially unexplained behavior* is a subsequence $S^{'}$ of S s.t. (i) $\mathscr{P}_P(S^{'}) \geq \tau$, (ii) $|S^{'}| \geq L$, and (iii) $S^{'}$ is minimal, i.e., there does not exist a subsequence $S^{''} \neq S^{'}$ of S s.t. $S^{''}$ is a subsequence of $S^{'}$, $\mathscr{P}_P(S^{''}) \geq \tau$, and $|S^{''}| \geq L$.

In the definition above, L is the minimum length a sequence must be for it to be considered a possible unexplained behavior. Totally unexplained behaviors (TUBs for short) $S^{'}$ have to be maximal because, based on Proposition 1, any subsequence of $S^{'}$ is totally unexplained with probability greater than or equal to that of $S^{'}$. On the other hand, partially unexplained behaviors (PUBs for short) $S^{'}$ have to be minimal because, based on Proposition 1, any super-sequence of $S^{'}$ is partially unexplained with probability greater than or equal to that of $S^{'}$.

Intuitively, an unexplained behavior is a sequence of events that are observed on a network and poorly explained by known behavior models. Such sequences might correspond to unknown variants of known behaviors or to entirely new – and unknown – behaviors. As such, the proposed approach may be help in discovering zero-day attacks, which are unknown to administrators by definition.

An *Unexplained Behavior Problem* (UBP) instance is a 4-tuple $I = \langle S, \mathscr{A}, \tau, L \rangle$ where S is an observation sequence, $\mathscr{A}$ is a set of behavior models, $\tau \in [0,1]$ is a probability threshold, and $L \in \mathbb{N}^+$ is a length threshold. We want to find the sets $\mathscr{O}^{tu}(I)$ and $\mathscr{O}^{pu}(I)$ of all totally and partially unexplained behaviors in S, respectively.

The unexplained behavior model presented in this section applies to both alert correlation and intrusion detection. When used at the intrusion detection level, observable events are packet features and models are IDS rules. When used at the alert correlation level, observable events are IDS alerts and models are attack models, such as attack graphs.

3 Properties

Previous work on the recognition of unexplained activities [3] relies on an independence assumption to break the large optimization problem of Definition 4 into smaller optimization problems[10]. Specifically, [3] uses the transitive closure of $\nsim$ to determine a partition of the set $\mathscr{O}$ of activity occurrences into equivalence classes $\mathscr{O}_1, \ldots, \mathscr{O}_m$, and assume that activity occurrences in one class are independent of activity occurrences in another class. Although this assumption is reasonable in the

[10] Indeed, the set of constraints becomes non-linear with the addition of the constraints reflecting the independence assumption.

realm of video data, where periods of low or no activity in the field of view of a single camera are likely to break the flow of events into independent segments, we drop such an assumption for the purpose of identifying unexplained behaviors in network intrusions. In fact, an observation stream typically includes alerts from multiple sources, and multiple activities may be occurring at any given time, making conflict based partitioning ineffective. For example, conflict based partitioning of the set of occurrences in Figure 3 leads to a single equivalence class containing all the occurrences.

In this section, we derive properties that can be leveraged to solve UBPs efficiently.

3.1 Totally unexplained behaviors

First, given a sequence $S^{'}$, we show that lower and upper bounds for $\mathscr{P}_T(S^{'})$ can be found without solving the optimization problem of Definition 4. In order to do so, we introduce the following preliminary definition.

Definition 6 (Maximal intersecting set of occurrences). Let $\mathscr{O}^*$ be a set of occurrences. A *maximal intersecting set of occurrences* for $\mathscr{O}^*$ is a subset $\mathscr{O}^{'}$ of $\mathscr{O}^*$ such that:

- $\forall o_i, o_j \in \mathscr{O}^{'}, o_i \nsim o_j$; and
- $\nexists \mathscr{O}^{''} \subseteq \mathscr{O}^*$ s.t. $\mathscr{O}^{'} \subset \mathscr{O}^{''} \wedge \forall o_i, o_j \in \mathscr{O}^{''}, o_i \nsim o_j$;

Intuitively, a set of occurrences is *intersecting* iff any two occurrences in the set conflict. An intersecting set of occurrences is *maximal* iff no proper superset of it is an intersecting set[11]. We use $\mathscr{M}(\mathscr{O}^*)$ to denote the set of maximal intersecting sets of occurrences in $\mathscr{O}^*$.

Example 6. Consider the observation sequence of Figure 3, and let $\mathscr{O}$ be the set of all occurrences recognized in the sequence. The set $\{o_4, o_5, o_6\}$ is a maximal intersecting set of occurrences for $\mathscr{O}$, as $o_4 \nsim o_5$, $o_4 \nsim o_6$, and $o_5 \nsim o_6$, and there is no proper superset containing pairwise conflicting occurrences. Instead, the set $\{o_3, o_4, o_5\}$ is not a maximal intersecting set of occurrences because o_3 and o_5 do not conflict. In this case, the set of all maximal intersecting sets of occurrences in $\mathscr{O}$ is $\mathscr{M}(\mathscr{O}) = \{\{o_1, o_2\}, \{o_2, o_3\}, \{o_1, o_2\}, \{o_3, o_4\}, \{o_4, o_5, o_6\}, \{o_6, o_7\}, \{o_7, o_8\}\}$.

Theorem 1. *Consider a subsequence $S^{'}$ of an observation sequence S and the set $\mathscr{O}$ of occurrences identified in S w.r.t. a set $\mathscr{A}$ of behavior models, and let $\mathscr{O}^*$ be the set of occurrences $o \in \mathscr{O}$ such that $o \cap S^{'} \neq \emptyset$. Then*

[11] The problem of finding all the maximal intersecting sets of occurrences is a generalization of the problem of finding maximal intersecting families of k-sets, but it is more general as occurrences are not required to have the same length k. As we need to compute maximal intersecting sets for small sets $\mathscr{O}^*$ of occurrences, complexity of this problem is not an issue.

$$\mathscr{P}_T(S^{'}) \geq 1 - \min\left\{1, \sum_{o \in \mathscr{O}^*} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)}\right\} \tag{1}$$

$$\mathscr{P}_T(S^{'}) \leq 1 - \max_{\mathscr{O}^{'} \in \mathscr{M}(\mathscr{O}^*)} \sum_{o \in \mathscr{O}^{'}} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)} \tag{2}$$

Proof. Consider a solution $[p_0, p_1, \ldots, p_n]^T$ of $\mathsf{LC}(S, \mathscr{A})$. Then $\mathscr{P}_T(S^{'}) = \sum_{w_i \in \mathscr{W} \; s.t. \; w_i \not\models_T S^{'}} p_i$. Recalling the definition of *totally unexplained sequence*, we can write

$$\mathscr{P}_T(S^{'}) = \sum_{w_i \in \mathscr{W} \; s.t. \; w_i \not\models_T S^{'}} p_i = \sum_{w_i \in \mathscr{W} \; s.t. \; \forall a_i \in S^{'}, \nexists o \in w_i, a_i \in o} p_i \tag{3}$$

Note that the condition $\forall a_i \in S^{'}, \nexists o \in w_i, a_i \in o$ is satisfied by all worlds except those containing at least one occurrence intersecting $S^{'}$. Therefore,

$$\sum_{w_i \in \mathscr{W} \; s.t. \; \forall a_i \in S^{'}, \nexists o \in w, a_i \in o} p_i = 1 - \sum_{w_i \in \mathscr{W} \; s.t. \; \exists o \in w_i, o \cap S^{'} \neq \emptyset} p_i \tag{4}$$

Lower bound. Recalling that $\mathscr{O}^*$ is the set of occurrences intersecting $S^{'}$, and considering that the condition $\exists o \in w_i, o \cap S^{'} \neq \emptyset$ is satisfied by all worlds w_i containing an occurrence $o \in \mathscr{O}^*$, with some worlds containing multiple such occurrences, we can write

$$\sum_{w_i \in \mathscr{W} \; s.t. \; \exists o \in w_i, o \cap S^{'} \neq \emptyset} p_i \leq \min\left\{1, \sum_{o \in \mathscr{O}^*} \sum_{w_i \in \mathscr{W} \; s.t. \; o \in w_i} p_i\right\} \tag{5}$$

Note that the argument $\sum_{w_i \in \mathscr{W} \; s.t. \; o \in w_i} p_i$ of the outer summation is the left-hand side of the constraint for occurrence o in the set of linear constraints of Definition 2. Therefore, combining Definition 2 and Equations 3, 4, and 5, we can write

$$\mathscr{P}_T(S^{'}) \geq 1 - \min\left\{1, \sum_{o \in \mathscr{O}^*} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)}\right\} \tag{6}$$

Upper bound. Consider a maximal intersecting set $\mathscr{O}^{'} \in \mathscr{M}(\mathscr{O}^*)$. For any two occurrences $o_i, o_j \in \mathscr{O}^{'}$, the sets of worlds $\mathscr{W}_i = \{w \in \mathscr{W} \mid o_i \in w\}$ and $\mathscr{W}_j = \{w \in \mathscr{W} \mid o_j \in w\}$ are disjoint, as $o_i \nsim o_j$. Additionally, the condition $\exists o \in w_i, o \cap S^{'} \neq \emptyset$ is satisfied in at least all worlds w_i containing an occurrence $o \in \mathscr{O}^{'}$, therefore,

$$\sum_{w_i \in \mathscr{W} \; s.t. \; \exists o \in w_i, o \cap S^{'} \neq \emptyset} p_i \geq \sum_{o \in \mathscr{O}^{'}} \sum_{w_i \in \mathscr{W} \; s.t. \; o \in w_i} p_i \tag{7}$$

As the above property holds for all $\mathscr{O}^{'} \in \mathscr{M}(\mathscr{O}^*)$, we can conclude that

$$\sum_{w_i \in \mathscr{W} \; s.t. \; \exists o \in w_i, o \cap S^{'} \neq \emptyset} p_i \geq \max_{\mathscr{O}^{'} \in \mathscr{M}(\mathscr{O}^*)} \sum_{o \in \mathscr{O}^{'}} \sum_{w_i \in \mathscr{W} \; s.t. \; o \in w_i} p_i \tag{8}$$

Finally, combining Definition 2 and Equations 3, 4, 8, we can write

$$\mathscr{P}_T(S') \leq 1 - \max_{\mathscr{O}' \in \mathscr{M}(\mathscr{O}^*)} \sum_{o \in \mathscr{O}'} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)} \tag{9}$$

□

Example 7. Consider again the observation sequence and occurrences of Figure 3. We want to find upper and lower bounds for the probability $\mathscr{P}_T(S')$ that the sequence $S' = \langle a_6, a_7, a_8, a_9, a_{10} \rangle$ is totally unexplained. For this example, $\mathscr{O}^* = \{o_3, o_4, o_5, o_6\}$ and $\mathscr{M}(\mathscr{O}^*) = \{\{o_3, o_4\}, \{o_4, o_5, o_6\}\}$. Applying Theorem 1 we obtain $\mathscr{P}_T(S') \geq 1 - 0.74 = 0.26$ and $\mathscr{P}_T(S') \leq 1 - \max\{0.36, 0.58\} = 0.42$. Note that, in this case, these bounds coincide exactly with the probability interval obtained by solving the maximization and minimization problems.

A consequence of Proposition 1 and Theorem 1 is the following theorem, which provides a sufficient condition for an observation not to be included in any unexplained behavior.

Theorem 2. *Let $I = \langle S, \mathscr{A}, \tau, L \rangle$ be a UBP instance. Given an observation $a \in S$, if $1 - \sum_{o \in \mathscr{O} \, s.t. \, a \in o} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)} < \tau$, then there does not exist a subsequence S' of S s.t. (i) $a \in S'$, (ii) $\mathscr{P}_T(S') \geq \tau$, and (iii) $|S'| \geq L$.*

Proof. Consider the sequence $S' = \langle a \rangle$. Then, $O^* = \{o \in \mathscr{O} \mid o \cap S' \neq \emptyset\} = \{o \in \mathscr{O} \mid a \in o\}$. Therefore, $\mathscr{M}(O^*) = \{O^*\}$, i.e., there is only one maximal intersecting set in O^*, and it coincides with O^*. Applying Theorem 1, we can conclude that

$$\mathscr{P}_T(S') \leq 1 - \sum_{o \in \mathscr{O}^*} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)} < \tau,$$

Now, consider a sequence $S'' \subseteq S$ s.t. $a \in S''$. Since $S' \subseteq S''$, from Proposition 1 we can conclude that

$$\mathscr{P}_T(S'') \leq \mathscr{P}_T(S') < \tau.$$

□

If the condition stated in the theorem above holds for an observation a, then we say that a is *sufficiently explained*. Note that checking whether an observation a is sufficiently explained does not require that we solve a set of linear constraints, since this can be done by simply summing the weighted probabilities of the occurrences containing a. Thus, this result yields a further efficiency. If a is sufficiently explained, then it can be disregarded for the purpose of identifying unexplained behaviors.

Given a UBP instance $I = \langle S, \mathscr{A}, \tau, L \rangle$ and a contiguous subsequence S' of S, we say that S' is a *candidate* iff (i) $|S'| \geq L$, (ii) $\forall a \in S'$, a is not sufficiently explained, and (iii) S' is maximal (i.e., there does not exist $S'' \neq S'$ s.t. S' is a subsequence of S'' and S'' satisfies (i) and (ii)). We use $candidates(I)$ to denote the set of candidate subsequences. If we look for totally unexplained behaviors that are contiguous

subsequences of S, then Theorem 2 entails that candidate subsequences can be individually considered because there is no (contiguous) totally unexplained behavior spanning two different candidate subsequences.

3.2 Partially unexplained behaviors

We now present similar results, in terms of probability bounds, for partially unexplained behaviors, and show that lower and upper bounds for $\mathscr{P}_P(S^{'})$ can be found without solving the optimization problem of Definition 4. In order to do so, we introduce the following preliminary definition.

Definition 7 (Non-conflicting sequence cover). Let $\mathscr{O}^*$ be a set of occurrences, and $S^{'}$ an observation sequence. A *non-conflicting sequence cover* of $S^{'}$ in $\mathscr{O}^*$ is a subset $\mathscr{O}^{'}$ of $\mathscr{O}^*$ such that:

- $\forall o_i, o_j \in \mathscr{O}^{'}, o_i$ and o_j do not conflict; and
- $\forall a \in S^{'}, \exists o \in \mathscr{O}^{'}, a \in o$.

We use $\mathscr{C}(\mathscr{O}^*, S^{'})$ to denote the set of all minimal sequence covers of $S^{'}$ in $\mathscr{O}^*$.

Intuitively, a *non-conflicting sequence cover* of $S^{'}$ in $\mathscr{O}^*$ is a subset of non-conflicting occurrences in $\mathscr{O}^*$ *covering*[12] $S^{'}$.

Theorem 3. *Consider a subsequence $S^{'}$ of an observation sequence S and the set O of occurrences identified in S, and let $\mathscr{O}^*$ be the set of occurrences $o \in \mathscr{O}$ such that $o \cap S^{'} \neq \emptyset$. Then*

$$\mathscr{P}_P(S^{'}) \geq 1 - \sum_{\mathscr{O}^{'} \in \mathscr{C}(\mathscr{O}^*, S^{'})} \min_{o \in \mathscr{O}^{'}} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)} \tag{10}$$

$$\mathscr{P}_P(S^{'}) \leq 1 - \sum_{o \in \mathscr{O}^* \, s.t. \, S^{'} \subseteq o} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)} \tag{11}$$

Proof. Consider a solution $[p_1, \ldots, p_n]^T$ of $\mathsf{LC}(S, \mathscr{A})$. Then $\mathscr{P}_P(S^{'}) = \sum_{w_i \in \mathscr{W} \, s.t. \, w_i \not\models_P S^{'}} p_i$. Recalling the definition of *partially unexplained sequence*, we can write

$$\mathscr{P}_P(S^{'}) = \sum_{w_i \in \mathscr{W} \, s.t. \, w_i \not\models_P S^{'}} p_i = \sum_{w_i \in \mathscr{W} \, s.t. \, \exists a_i \in S^{'}, \nexists o \in w_i, a_i \in o} p_i \tag{12}$$

Note that the condition $\exists a_i \in S^{'}, \nexists o \in w_i, a_i \in o$ is satisfied by all worlds except those where each observation $a_i \in S^{'}$ is part of an occurrence, that is the sequence is *totally explained* in those worlds. Therefore,

[12] This is a variant of the set cover problem. This is known to be NP-complete, however we need to solve only small instances of this problem, so complexity is not an issue.

$$\sum_{w_i \in \mathscr{W} \, s.t. \, \exists a_i \in S^{'}, \nexists o \in w, a_i \in o} p_i = 1 - \sum_{w_i \in \mathscr{W} \, s.t. \, \forall a_i \in S^{'}, \exists o \in w_i, a_i \in o} p_i \tag{13}$$

Lower bound. Given any two non-conflicting sequence covers $\mathscr{O}^{'}$ and $\mathscr{O}^{''}$, the sets of worlds $\mathscr{W}^{'} = \{w \in \mathscr{W} \mid \mathscr{O}^{'} \subseteq w\}$ and $\mathscr{W}^{''} = \{w \in \mathscr{W} \mid \mathscr{O}^{''} \in w\}$ are disjoint, as at least one occurrence in $\mathscr{O}^{'}$ conflicts with at least one occurrence in $\mathscr{O}^{''}$. Additionally, the condition $\forall a_i \in S^{'}, \exists o \in w_i, a_i \in o$ is satisfied by all worlds w_i containing a non-conflicting cover of $S^{'}$. Thus, we can write

$$\sum_{w_i \in \mathscr{W} \, s.t. \, \forall a_i \in S^{'}, \exists o \in w_i, a_i \in o} p_i = \sum_{\mathscr{O}^{'} \in \mathscr{C}(\mathscr{O}^*, S^{'})} \sum_{w_i \in \mathscr{W} \, s.t. \, \mathscr{O}^{'} \subseteq w_i} p_i \tag{14}$$

Consider any non-conflicting sequence cover $\mathscr{O}^{'} \in \mathscr{C}(\mathscr{O}^*, S^{'})$. The set $\mathscr{W}^{'} = \{w_i \in \mathscr{W} \; s.t. \; \mathscr{O}^{'} \subseteq w_i\}$ of worlds containing all the occurrences in $\mathscr{O}^{'}$ is a subset of the set of worlds containing o, for each $o \in \mathscr{O}^{'}$. Therefore,

$$\sum_{w_i \in \mathscr{W} \; s.t. \; \mathscr{O}^{'} \subseteq w_i} p_i \leq \min_{o \in \mathscr{O}^{'}} \sum_{w_i \in \mathscr{W} \; s.t. \; o \in w_i} p_i \tag{15}$$

Finally, considering that the above property holds for any $\mathscr{O}^{'} \in \mathscr{C}(\mathscr{O}^*, S^{'})$, and combining Definition 2 and Equations 12, 13, 14 and 15, we can write

$$\mathscr{P}_P(S^{'}) \geq 1 - \sum_{\mathscr{O}^{'} \in \mathscr{C}(\mathscr{O}^*, S^{'})} \min_{o \in \mathscr{O}^{'}} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)}$$

Upper bound. Consider the set $\mathscr{O}^{'} = \{o \in \mathscr{O}^* \, s.t. \, S^{'} \subseteq o\}$ of all the occurrences o that cover $S^{'}$, i.e., the occurrences such that $\{o\}$ is a sequence cover for $S^{'}$ in $\mathscr{O}^*$. $S^{'}$ is *totally explained* in at least all the worlds containing any of the occurrences in $\mathscr{O}^{'}$. Note that any two set of worlds W_i and W_j containing $o_i \in \mathscr{O}^{'}$ and $o_j \in \mathscr{O}^{'}$ respectively are disjoint, as $o_i \nsim o_j$. Therefore,

$$\sum_{w_i \in \mathscr{W} \, s.t. \, \forall a_i \in S^{'}, \exists o \in w_i, a_i \in o} p_i \geq \sum_{o \in \mathscr{O}^* \, s.t. \, S^{'} \subseteq o} \sum_{w_i \in \mathscr{W} \; s.t. \; o \in w_i} p_i \tag{16}$$

Finally, combining Definition 2 and Equations 12, 13 and 16, we can write

$$\mathscr{P}_P(S^{'}) \leq 1 - \sum_{o \in \mathscr{O}^* \, s.t. \, S^{'} \subseteq o} p^*(o) \cdot \frac{\omega(o)}{\sum_{o_j \in C(o)} \omega(o_j)}$$

□

Example 8. Consider again the observation sequence and occurrences of Figure 3. We want to find upper and lower bounds for the probability $\mathscr{P}_P(S^{''})$ that the sequence $S^{''} = \langle a_7, a_8 \rangle$ is partially unexplained. For this example, $\mathscr{C}(\mathscr{O}^*, S^{''}) = \{\{o_3, o_5\}, \{o_4\}\}$. Applying Theorem 3 we obtain $\mathscr{P}_P(S^{''}) \geq 1 - 0.36 = 0.64$ and $\mathscr{P}_P(S^{''}) \leq 1 - 0.2 = 0.8$. Note that, in this case, these bounds coincide exactly with the probability interval obtained by solving the maximization and minimization problems.

4 Algorithms

Even though our framework can assess the probability that an arbitrary subsequence of S is unexplained, we propose algorithms that search for contiguous subsequences of S, as we believe that contiguous subsequence can more easily be interpreted by users (nevertheless, the algorithms could be easily modified to identify also non-contiguous unexplained subsequences).

We now present algorithms to find totally and partially unexplained behaviors. These algorithms are a variant of the algorithms in [3]: while the algorithms in [3] compute the *exact* probability that a sequence is unexplained, the algorithms in this paper compute an *approximate* probability that a sequence is unexplained by leveraging the properties shown in Section 3.

Given an observation sequence $S = \langle a_1, \ldots, a_n \rangle$, we use $S(i, j)$ $(1 \leq i \leq j \leq n)$ to denote the subsequence $S' = \langle a_i, \ldots, a_j \rangle$.

The FindTUB algorithm computes totally unexplained behaviors in an observation sequence S. Leveraging Theorem 2, FindTUB only considers candidate subsequences of S. When the algorithm finds a sequence $S'(start, end)$ of length at least L having a probability of being unexplained greater than or equal to τ (line 6), then the algorithm makes it maximal by adding observations on the right. Instead of adding one observation at a time, $S'(start, end)$ is extended of L observations at a time until its probability drops below τ (lines 8–13); then, the exact maximum length of the unexplained behavior is found by performing a binary search between s and e (line 16). Note that $\mathscr{P}_T$ is estimated by applying Theorem 1.

The FindPUB algorithm computes all partially unexplained behaviors. To find an unexplained behavior, it starts with a sequence of a certain length (at least L) and adds observations on the right of the sequence until its probability of being unexplained is greater than or equal to τ. As in the case of FindTUB, this is done not by adding one observation at a time, but adding L observations at a time (lines 7–11) and then determining the exact minimal length by performing a binary search between s and e (line 16). The sequence is then shortened on the left making it minimal by performing a binary search instead of proceeding one observation at a time (line 23). Note that $\mathscr{P}_P$ is estimated by leveraging Theorem 3.

5 Experimental Results

In this section, we present the results of experiments we conducted on a prototype implementation of the proposed framework. We evaluated running time of the algorithms as well as accuracy. In the following, we first describe the experimental setup (Section 5.1), and then report the results on the scalability (Section 5.2) and accuracy (Section 5.3) of our framework.

Algorithm 1 FindTUB(I)

Require: UBP instance $I = \langle S, \mathscr{A}, \tau, L \rangle$
Ensure: Set $\mathscr{O}^{tu}$ of totally unexplained behaviors

$\mathscr{O}^{tu} = \emptyset$
for all $S' \in relevant(I)$ **do**
  $start = 1$
  $end = L$
  repeat
    if $\mathscr{P}_T(S'(start, end)) \geq \tau$ **then**
      $end' = end$
      while $end < |S'|$ **do**
        $end = \min\{end + L, |S'|\}$
        if $\mathscr{P}_T(S'(start, end)) < \tau$ **then**
          break
        end if
      end while
      $s = \max\{end - L, end'\}$
      $e = end$
      $end = \max\{mid \in \mathbb{N}^+ \mid s \leq mid \leq e \wedge \mathscr{P}_T(S(start, mid)) \geq \tau\}$
      $S'' = S'(start, end)$
      Add S'' to $\mathscr{O}^{tu}$
      $start = start + 1$
      $end = start + |S''| - 1$
    else
      $start = start + 1$
      $end = \max\{end, start + L - 1\}$
    end if
  until $end > |S'|$
end for
return $\mathscr{O}^{tu}$

5.1 Experimental Setup

All experiments were conducted on a dataset consisting of network traffic captured over a 24-hour period from the internal network of an academic institution. We used (i) Wireshark (`http://www.wireshark.org/`) to capture network traffic and generate the sequence of packets, and (ii) Snort (`http://www.snort.org/`) to analyze such traffic and generate the sequence of alerts.

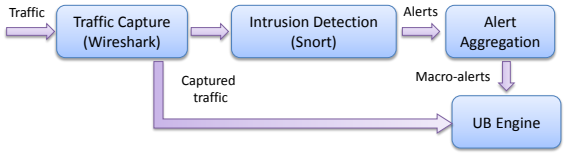


Fig. 5 Experimental setup

Algorithm 2 FindPUB(I)

Require: UBP instance $I = \langle S, \mathscr{A}, \tau, L \rangle$
Ensure: Set $\mathscr{O}^{pu}$ of partially unexplained behaviors

$\mathscr{O}^{pu} = \emptyset$
$start = 1$
$end = L$
while $end \leq |S|$ **do**
  if $\mathscr{P}_P(S(start, end)) < \tau$ **then**
    $end' = end$
    while $end < |S|$ **do**
      $end = \min\{end + L, |S|\}$
      if $\mathscr{P}_P(S(start, end)) \geq \tau$ **then**
        break
      end if
    end while
    if $\mathscr{P}_P(S(start, end)) \geq \tau$ **then**
      $s = \max\{end' + 1, end - L + 1\}$;
      $e = end$
      $end = \min\{mid \in \mathbb{N}^+ \mid s \leq mid \leq e \wedge \mathscr{P}_P(S(start, mid)) \geq \tau\}$
    else
      return $\mathscr{O}^{pu}$
    end if
  end if
  $s' = start$
  $e' = end - L + 1$
  $start = \max\{mid \in \mathbb{N}^+ \mid s' \leq mid \leq e' \wedge \mathscr{P}_P(S(mid, end)) \geq \tau\}$
  $S' = S(start, end)$
  Add S' to $\mathscr{O}^{pu}$
  $start = start + 1$
  $end = start + |S'| - 1$
end while
return $\mathscr{O}^{pu}$

Figure 5 illustrates the experimental setup. As the number of alerts returned by the IDS may be relatively high, the *Alert Aggregation* module, that takes as input the identified alerts, can optionally aggregate multiple alerts triggered by the same event into a macro-alert, based on a set of ad hoc aggregation rules. For instance, we defined rules to aggregate alerts such that protocol, source address, and destination address of suspicious traffic are the same, and the alerts are within a given temporal window. In other words, the events triggering such alerts will be treated as a single event, thus reducing the amount of data to be processed.

5.2 Scalability Results

We measured the running time of FindTUB for different values of τ and L, varying the length of the data stream to be analyzed. More specifically, in one case we set the threshold τ to 0.6 and used different values of L, namely $10, 50, 75, 100$. In the other case, we fixed the value of L to 50 and varied the threshold τ giving it the values $0.4, 0.6, 0.8$.

Figure 6 shows the processing time of FindTUB as a function of the data stream length (expressed in seconds) for different values of L. Not surprisingly, the running time increases as the input stream size grows. Moreover, the processing time decreases as the value of L increases because Algorithm FindTUB can move forward in the data stream more quickly for higher values of L. Notice also that the running times is much lower when $L \geq 50$.

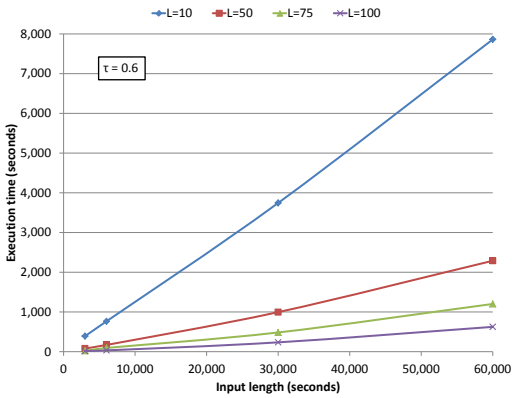


Fig. 6 Execution time ($\tau = 0.6$, $L = 10, 50, 75, 100$)

Figure 7 shows the processing time of FindTUB as a function of the data stream length for different values of τ. Also in this case the running time gets higher as the input stream length increases. The running time is lower for higher values of τ because the pruning strategy of Algorithm FindTUB becomes more effective with higher threshold values. Moreover, in this case, the running time becomes much lower when $\tau \geq 0.6$.

Both figures show that our algorithm scales well – notice that the running time linearly grows w.r.t. the length of the input.

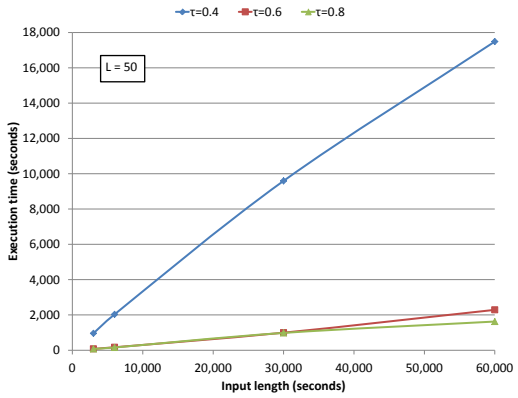


Fig. 7 Execution time ($L = 50$, $\tau = 0.4, 0.6, 0.8$)

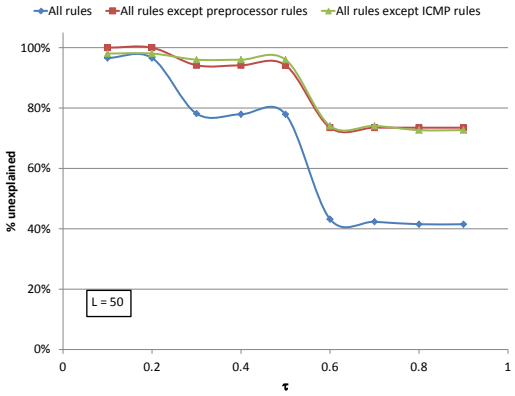
5.3 Accuracy Results

We measured the accuracy of the framework using the following procedure. Let $\mathscr{A}$ be the set of Snort rules. First, we detected all occurrences of $\mathscr{A}$ in the considered data stream. Then, we executed multiple runs of FindTUB, and at each run i, we ignored a different subset $\mathscr{A}_i$ of $\mathscr{A}$. Clearly, ignoring models in $\mathscr{A}_i$ is equivalent to not having those models available. Thus, occurrences of ignored behaviors are expected to have a relatively high probability of being unexplained as there is no model for them. We measured the fraction of such occurrences that have been flagged as unexplained by FindTUB for different values of τ, namely $0.4, 0.6, 0.8$ (L was set to 50).

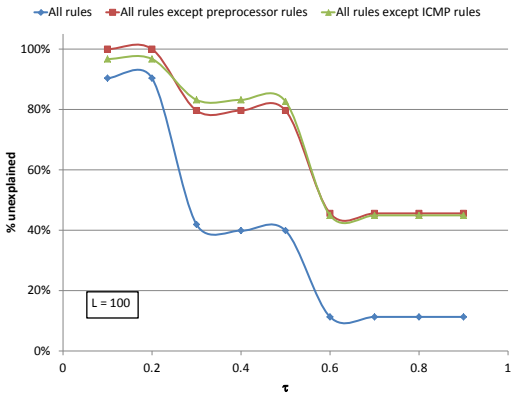
We considered two settings: one where only *ICMP rules* in $\mathscr{A}$ were ignored, and another one where only *preprocessor rules* in $\mathscr{A}$ were ignored. The average accuracy in the former and latter case is shown in Table 2 and Table 3, respectively. Notice that the accuracy decreases as the threshold value increases since higher thresholds are more restrictive conditions for a sequence to be unexplained. Notice also that in both cases there is no difference in accuracy between $\tau = 0.6$ and $\tau = 0.8$; this is because, in this case, the same unexplained sequences were found. These results show that our framework achieved high accuracy.

τ	Accuracy
0.4	95.10%
0.6	78.75%
0.8	78.75%

Table 2 Accuracy when ICMP rules are ignored



(a) Unexplained traffic for $L = 50$



(b) Unexplained traffic for $L = 100$

Fig. 8 Percentage of unexplained traffic vs. τ for different values of L and different sets of rules

τ	Accuracy
0.4	84.21 %
0.6	73.68 %
0.8	73.68 %

Table 3 Accuracy when preprocessor rules are ignored

We also evaluated the effect of the threshold value on the percentage of unexplained traffic. We considered three different settings, each characterized by a different set of available rules (all Snort rules are available, all Snort rules except for prepocessor rules are available, and all rules but ICMP rules are considered), and measured the percentage of unexplained traffic as the threshold varies from 0.1 to 1. We carried out the experiments for two different values of L, namely $L = 50$

and $L = 100$ – the results are reported in Figure 8(a) and Figure 8(b), respectively. Both figures show that by disregarding rules the percentage of unexplained traffic increases, but there is no substantial difference between disregarding preprocessor rules and disregarding ICMP rules. Furthermore, the percentage of unexplained traffic decreases as the threshold increase because higher threshold values impose more restrictive conditions for a sequence to be unexplained. Finally, the results for $L = 100$ show lower percentages of unexplained traffic than the case $L = 50$ as $L = 100$ is a more restrictive condition for a sequence to be unexplained and thus the unexplained traffic is expected to be less in this case. This trend is also confirmed by the results of Figure 9 where we show how the percentage of unexplained traffic varies as the threshold value goes from 0.1 to 1 and different values of L are considered. In this case L was set to $10, 50, 100, 200$ and all IDS rules were considered.

6 Conclusions

In this paper, we presented a probabilistic framework to identify unexplained behavior in network intrusions. Intrusion detection and alert correlation methods rely on models encoding a priori knowledge of either normal or malicious behavior, but are incapable of quantifying how well the underlying models explain what is observed on the network. Our framework addresses this limitation, by evaluating the probability that a sequence of events is unexplained, given a set of models. We derived some important properties of the framework that can be leveraged to estimate this probability efficiently. The proposed framework can operate both at the intrusion detection level and at the alert correlation level. Experimental results show that the algorithms are accurate and scale linearly with the size of the observation se-

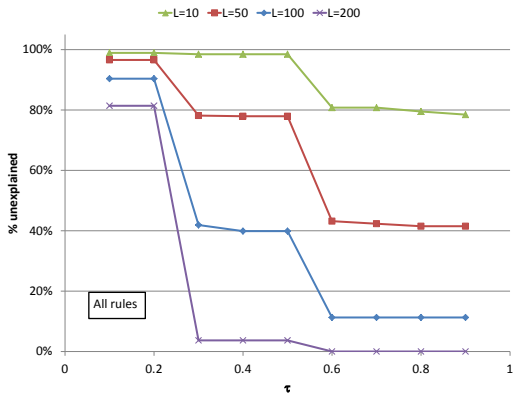


Fig. 9 Percentage of unexplained traffic vs. τ for different values of L (all IDS rules are used)

quence. This confirms the validity of our approach and motivate further research in this direction.

References

1. S. O. Al-Mamory and H. Zhang. Ids alerts correlation using grammar-based approach. *Journal of Computer Virology*, 5(4):271–282, November 2009.
2. M. Albanese, S. Jajodia, A. Pugliese, and V. S. Subrahmanian. Scalable analysis of attack scenarios. In *Proceedings of the 16th European Symposium on Research in Computer Security (ESORICS 2011)*, pages 416–433, Leuven, Belgium, September 2011. Springer.
3. M. Albanese, C. Molinaro, F. Persia, A. Picariello, and V. S. Subrahmanian. Finding "unexplained" activities in video. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, pages 1628–1634, Barcelona, Spain, July 2011.
4. J. P. Anderson. Computer security threat monitoring and surveillance. Technical report, James P. Anderson Co., Fort Washington, PA, USA, April 1980.
5. H. Debar and A. Wespi. Aggregation and correlation of intrusion-detection alerts. In W. Lee, L. Mé, and A. Wespi, editors, *Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection (RAID 2001)*, volume 2212 of *Lecture Notes in Computer Science*, pages 85–103, Davis, CA, USA, October 2001. Springer.
6. P. García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández, and E. Vázquez. Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, 28(1-2):18–28, February-March 2009.
7. A. Jones and S. Li. Temporal signatures for intrusion detection. In *Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC 2001)*, pages 252–261, New Orleans, LA, USA, December 2001. IEEE Computer Society.
8. B. Mukherjee, L. T. Heberlein, and K. N. Levitt. Network intrusion detection. *IEEE Network*, 8(3):26–41, May 1994.
9. P. Ning, Y. Cui, and D. S. Reeves. Constructing attack scenarios through correlation of intrusion alerts. In *Proceedings of the 9th ACM Conference on Computer and Communications Security(CCS 2002)*, pages 245–254, Washington, DC, USA, November 2002. ACM.
10. S. Noel, E. Robertson, and S. Jajodia. Correlating intrusion events and building attack scenarios through attack graph distances. In *Proceedings of the 20th Annual Computer Security Applications Conference (ACSAC 2004)*, pages 350–359, Tucson, AZ, USA, December 2004.
11. A. J. Oliner, A. V. Kulkarni, and A. Aiken. Community epidemic detection using time-correlated anomalies. In S. Jha, R. Sommer, and C. Kreibich, editors, *Proceedings of the 13th International Symposium on Recent Advances in Intrusion Detection (RAID 2010)*, volume 6307 of *Lecture Notes in Computer Science*, pages 360–381, Ottawa, Canada, September 2010. Springer.
12. X. Qin. *A Probabilistic-Based Framework for INFOSEC Alert Correlation*. Phd thesis, Georgia Institute of Technology, August 2005.
13. X. Qin and W. Lee. Statistical causality analysis of INFOSEC alert data. In G. Vigna, C. Kruegel, and E. Jonsson, editors, *Proceedings of the 6th International Symposium on Recent Advances in Intrusion Detection (RAID 2003)*, volume 2820 of *Lecture Notes in Computer Science*, pages 73–93, Pittsburgh, PA, USA, September 2003. Springer.
14. O. Sheyner, J. Haines, S. Jha, R. Lippmann, and J. M. Wing. Automated generation and analysis of attack graphs. In *Proceedings of the 2002 IEEE Symposium on Security and Privacy (S&P 2002)*, pages 273–284, Berkeley, CA, USA, May 2002.
15. L. Wang, A. Liu, and S. Jajodia. Using attack graphs for correlating, hypothesizing, and predicting intrusion alerts. *Computer Communications*, 29(15):2917–2933, September 2006.