

## Correlation of IDS Events

by: Ramesh Sripathy Rao & Elango Krishnasami, 02/23/2005

<http://www.securitydocs.com/library/3030>

*Note : This paper has been reformatted and republished from our archives.*

### Abstract

Recently there have been much interest in Event correlation to computer network intrusion detection events to speculate the pattern of an attack. This paper explores some correlation techniques which can be applied to the Intrusion alerts and identify the patterns that are seen commonly across the events.

### Introduction

Intrusion detection starts with instrumentation of a computer network for data collection. Signature-based software 'sensors(a hardware or a software that looks at all the traffic for known attacks)' monitor the network traffic and raise 'alarms' when the traffic matches specified attack signatures. Security analysts look at each event and decide whether the alarm indicates an event serious enough to demand attention. A response might be to block the traffic from specific address or network or to call the internet service provider associated with suspicious traffic, or to simply make note of unusual traffic for future reference/baseline.

If the network is small and signatures are kept up to date, the human analyst solution for Intrusion detection works well. But when organizations have a large, complex network, the number of alarms they need to review overwhelms human analysts.

This situation arise from ever increasing attacks on the network, as well as the tendency of sensor signature patterns to be insufficiently selective (i.e., raise too many false alarms).

Many commercial tools provide console for real time viewing of events and reports to do analysis. But real-time viewing might not be helpful for doing analysis, analyst might not be able to correlate events happened across the networks. The reports provided by most of the vendors are presented in textual format and needs user inputs, which might not be of much help to the security analysts.

The purpose of this paper is to suggest some methods for analyzing events automatically, which could help the security analysts for identifying common patterns that are seen across the events.

The approach discussed below will try to analyze the events and help the user in identifying a pattern, which has been seen always with certain attacks. This will help the analyst to predict when some events (of low priority) occur in the network, and can be followed by more serious attacks later.

### Correlating attacks spanning multiple destinations

When an attack is seen for a specific host in a network the same attack might have occurred for other hosts in different or in the same network. The attacker might have used the similar pattern for generating the attacks. This method will try to identify those common pattern or attacks that were followed by the attackers before generating a serious attack. This method summarizes this information to the user so that

the user can configure his Security Policy for a particular pattern of event.

This method will try to identify patterns, which are seen across multiple destinations and display the events, which occurs most of the times.

The security analyst will configure signatures, which is more important for his network with a higher severity. This method will select those high severity signatures for which events are seen and try to see if the same high severity event has occurred across multiple destinations in his other networks. If it has occurred across multiple destinations then we identify whether we see any common events among these destinations, which has occurred before this attack. With this the system would be able to list those common events that have been followed by an attacker before a serious attack. This would help the security analyst in identifying the pattern followed by the attackers and change his security policy.

### Algorithm

For each high severity event S0,

- Let F0 be a set of all destinations that has S0.
- Let F1 be a set of events occurred before S0 on each of the destinations in F0.
- For each event in F1 find the weight. Where weight=no. of occurrences of the event in F0.
- Let T1 be the set of top N weights and C1 be set of corresponding confidence (formula (weight/no of destinations in F0) \* 100).
- Show that S0 would happen if T1 happens with the confidence level C1.

Let's take some events as an example for understanding the algorithm.

Severity	Source Address	Destination Address	Attack Type
High	Attacker1	Victim1	ABC
High	Attacker2	Victim2	ABC
High	Attacker1	Victim3	ABC
Low	Attacker3	Victim1	XYZ
Low	Attacker	Victim2	XYZ1
Low	Attacker	Victim3	XYZ2
Low	Attacker3	Victim1	XYZ
Low	Attacker	Victim1	XYZ
Low	Attacker	Victim2	XYZ
Low	Attacker	Victim3	XYZ1
Low	Attacker	Victim1	XYZ2

Low	Attacker	Victim1	XYZ
Low	Attacker	Victim2	XYZ2
Low	Attacker	Victim3	XYZ

From the events above “ABC” is a high severity signature seen across many victims. Let event type “ABC” is S0. Let F0 is the set of all destinations that has S0. Victim 1, Victim 2, Victim 3 all has S0. Events “XYZ, XYZ1, XYZ2” are all seen before S0 for all the Victims. Let F1 be these events.

Weights for F1

Event “XYZ” occurred 3 times. This event is seen for all the victims.

Event “XYZ1” occurred 2 times. This event is seen for Victim 2 and Victim 3.

Event “XYZ2” occurred 2 times. This event is seen for only Victim 2 and Victim 3.

T1

“XYZ, XYZ2, XYZ1” in the order.

Confidence of the events followed before S0 = ((weight/no. of destinations in F0) \*100)

For “XYZ”

$$C = (3/3) * 100 = 100 \%$$

For “XYZ1”

$$C1 = (2/3) * 100 = 66\%$$

For “XYZ2”

$$C2 = (2/3) * 100 = 66 \%$$

Now we can conclude to the user that When S0 (ABC) could happen if XYZ (100 %) or XYZ1 (66%) or XYZ2 (66%) is happened.

### Correlating attacks targeted to single host

There are some attacks which might be targeted to some specific hosts in a network. The attackers might follow certain patterns while launching the attacks. This method will try to identify attacks occurred for some specific hosts and try to look for patterns or attacks that was done by the attacker before launching a more serious attack.

This method will allow us to identify attacks which are targeted to single host unlike the first method which will try to identify attacks occurring across multiple hosts and identify patterns which seen most of the times. We identify a high severity event and then split the complete data into pieces of time interval and then try to identify if the same event has occurred to the same host in different time window, if so we try identifying events, which occurred before this high severity event. Among these we select the events that are common across and will be able to predict this attack for that specific host.

### Algorithm

For each high severity event S0.

- Let F0 be a destination that has S0.
- Select events by splitting the whole data into multiples of "w" time window.
- For each of these time window
  - Let F1 be list of all such events that has occurred to the same destination F0 but before S0.

- For each event in F1 find the weight. Where weight=no. of occurrences of the event across different time window in “w”.
- Let T1 be the set of top N weights and C1 be the confidence (formula (weight/no of time intervals with F1) \* 100)
- Show that S0 would happen if T1 happens with the confidence level C1.

Severity	Source Address	Destination Address	Attack Type	Time
High	Attacker1	Victim1	ABC	
High	Attacker2	Victim1	CDE	
Low	Attacker3	Victim1	XYZ	
Low	Attacker	Victim1	XYZ1	
Low	Attacker	Victim1	XYZ2	
< different time Interval >				
High	Attacker1	Victim1	ABC	
High	Attacker2	Victim1	CDE	
Low	Attacker	Victim1	XYZ1	
Low	Attacker3	Victim1	XYZ	
Low	Attacker	Victim1	XYZ2	
< different time Interval >				
High	Attacker1	Victim1	ABC	
High	Attacker2	Victim1	ABC1	
Low	Attacker3	Victim1	XYZ	
Low	Attacker	Victim1	XYZ11	
Low	Attacker	Victim1	XYZ21	
< different time Interval >				
High	Attacker1	Victim1	ABC	
High	Attacker2	Victim1	CDE	
Low	Attacker3	Victim1	XYZ	
Low	Attacker	Victim1	XYZ11	
Low	Attacker	Victim1	XYZ2	

Sample events for explanation

From the example let's take the high severity event "ABC" as S0.

Now let us split the whole event data in to different time window of "W".

In each of the time window "W" look for event "ABC" for the same victim.

Let "XYZ, XYZ11, XYZ2, and XYZ21" be F1, events that has occurred before S0.

Let us calculate the weights for each of the event in F1 across all the time interval "W".

Weight for "XYZ" is 4.

Weight for "XYZ11" is 2.

Weight for "XYZ2" is 3.

Weight for "XYZ21" is 1.

Top "N" weights from the example are "XYZ, XYZ2, and XYZ21".

Now to calculate the Confidence of the events followed before S0

Confidence = ((weight/ no of time intervals with F1) \*100)

No .of time intervals in the example are "4".

For "XYZ"

$$C = (4/4)*100 = 100\%$$

For "XYZ2"

$$C = (3/4)*100 = 75\%$$

For "XYZ11"

$$C = (2/4)*100 = 50\%$$

Now we can conclude to the user that When S0 (ABC) could happen if XYZ (100 %) or XYZ2 (75%) or XYZ11 (50%) is happened

This method will try to identify events that are targeted for a specific host rather than multiple hosts as the method 1.

**Disadvantages:**

If unrelated event is seen across different time intervals or across the different hosts then the method discussed above might report that event as correlated even if they are not related.

**Correlating attacks based on Time.**

This method will allow us to identify attacks which are happening regularly at the some specified time. This method finds out the attack patterns based on the timing of the attack. This analysis will give an overview on the attacks, which are happening over the weekend, working hours and non working hours. Based on the types of attacks and timing the security policy can be strengthened for different timing.

The attack patterns are identified by sorting the events based on the time and by finding the common attacks at the specified time interval over the period of days. All info and Low severity event might lead to other severe attacks. So even though analyzing these attacks requires huge memory & time consuming process, all the low and info events will be included during the analysis.

Sort the events occurred based on the event type or event id. The sorted events are filtered based the on the time interval over the period of time say 'N' number of days. Take each event type and look for any

common patterns followed on the given time interval, say every day or every week. Event occurrences are calculated and the confidence level is calculated by using no of occurrences and no of days of analysis. Events those are not repeated again or don't follow any timing patterns will be ignored.

### Algorithm

Let E0 be the event name

- For each Time Interval T0
- Let D0 be the number of days event was taken for mining
- Check the E0 exists on interval T0 for D0 days
- N0 be the total number of occurrences for the time interval T0
- Repeat the same for different time interval T1, T2,...,Tn
- Sort the Nn (number of occurrences) respect to Tn
- Let H1, H2 & H3 be the top three counts
- List H1, H2 & H3 with respective Tn
- Confidence level C = No of occurrences/No of days

Repeat the procedure for all occurred Events

This method is based on the event type and the time of the event occurred. This can be extended to include victim based to find out what sort of attacks coming at the specified time interval for the specified victim host. That will help the security analyst to update the security policy for the key servers.

Sample events for explanation

Severity	Attack Type	Day	Time
Medium	ABC	Day1	7 AM
Medium	ABC	Day1	8 AM
Medium	CDE	Day1	9 AM
Medium	ABC	Day1	6 PM
Medium	CDE	Day1	7 PM
Medium	ABC	Day2	7:30 AM
Medium	ABC	Day2	8:15 AM
Medium	CDE	Day2	1 PM
Medium	ABC	Day2	9 PM
Medium	CDE	Day2	7 PM
Medium	ABC	Day3	1 AM
Medium	ABC	Day3	1 AM
Medium	ABC	Day3	6:30 AM
Medium	CDE	Day3	6 PM
Medium	ABC	Day3	4 PM

Medium	ABC	Day3	7 PM
Medium	CDE	Day3	11 PM

From the above example, take the event ABC and assume four time intervals  
T0 (6AM to 9 PM), T1 (9 AM to 6 PM), T2 (6 PM to 9 PM) & T3 (9 PM to 6 AM)  
D0 is the 3 days of event

N0 is 3 for T0 with confidence level of 100%  
N1 is 2 for T1 with confidence level of 66%  
N2 is 1 for T2 with confidence level of 33%  
N3 is 1 for T3 with confidence level of 33%

This data shows the Event ABC Regularly occurs in the time interval T0 & Frequently occurs in the time interval T1 with confidence level of 66%

## Conclusion

In this document we saw the challenges faced by the security analyst in analyzing the events generated by the IDS as well as approaches that can be taken for doing analysis of IDS events. Although the above mentioned techniques doesn't provide a guaranteed and full fledged solution, it sure does give an insight on how to proceed in correlating alarms in a determined fashion.

## References:

1. [www.snort.org](http://www.snort.org)
2. [www.insecure.org](http://www.insecure.org)
3. [www.sans.org](http://www.sans.org)

## Author Biography:

Ramesh Sripathy Rao, M.C.A., M.S (Software Engg.), has been associated with Cisco Offshore Development Center (HCL Technologies Ltd., India) since 1999. Reachable at [sramesh\\_11@yahoo.com](mailto:sramesh_11@yahoo.com).

Elango Krishnasami B.E., has been associated with Cisco Offshore Development Center (HCL Technologies Ltd., India) since 1999. Reachable at [elangok@yahoo.com](mailto:elangok@yahoo.com).