Event normalization in SIEM

Each log source have own event format. In enterprise environment there can be hundreds of systems from which we collect and analyze logs. Ideally analyst must know format for each system or can quickly dive into them. Each system have own purpose and events can have different information. Assume we start event collection and incident occuried. How we can find events that relate to incident? Which query we must run to find this events?

Preface

Normalization provide ablitity to map event fields from any log source to standard scheme or framework. Each log source type can have own log format and content. Web proxy logs contains source IP-address URL, status code, browser name and version and etc. Antispam logs contains sender and destination email addresses, source IP-address, source domain, spam score. Firewall logs contains source and destination IP-addresses and ports, protocol and etc. In this examples there the same event fields like source IP-address and a lot of different. To support huge number of systems normalization scheme must take into consideration on them.

Normalization scheme

Number of elements in scheme is a long time dispute. There 2 main approaches when we examine normalization scheme:

1. Short scheme. Only most popular fields are used. About 20 fields. For example, username, source and destination IP-addresses and ports and etc. This fields supported for any supported log source. Analyst can add custom fields to scheme. Sometimes custom fields cannot be used as variables, indexes and etc. This is limitation for custom fields. For big installations this is serious limitations. The value of short scheme – good performance and simplicity. Unfortunately any new custom field for each log source type requires time resources for writing parsers and mapping.
2. Full scheme. 99.9% of common fields are included in this scheme. About 300-400 fields. There few reserved custom fields to which analyst can map exotic fields from events.

How normalization work

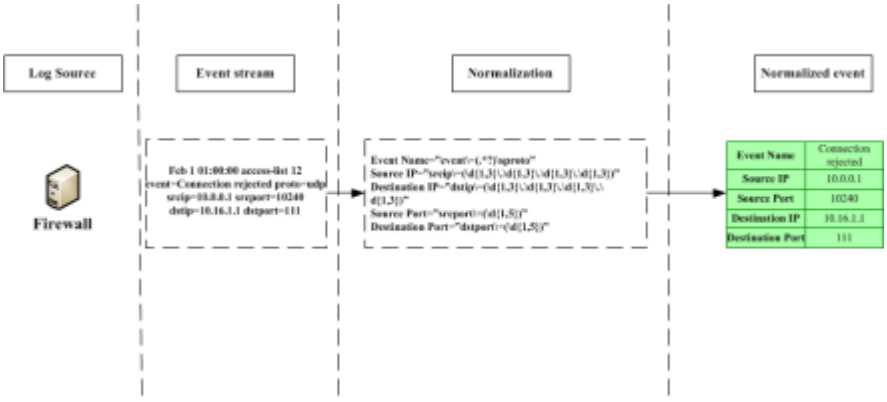
If we does not deep into technical details:

1. Raw event stream received by collector/connector.
2. Collector find to which log source type besides event and load parser or take it from cache.
3. For each event applied parser. Parser is set of regex. Each regex used to find field(source ip, destination port, username and etc.) in event.
4. Event normalized and categorized.
5. Aggregation and filtering applied.
6. Next event coming.

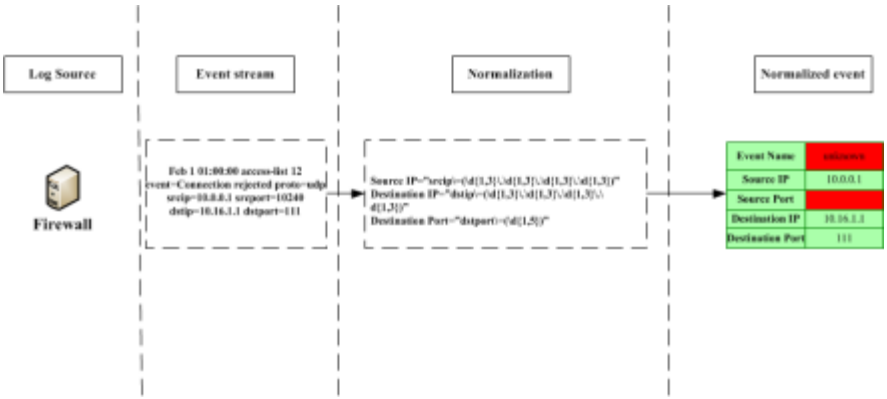
Check normalization in SIEM

Quality of parser and normalization depends on developer. Quality – percentage of events and fields from event which will be normalized. More percentage = better quality. A lot of systems exists and must be supported. Writing good parsers consume time and human resources. Some vendors can select only critical security events for normalization. If we multiple it on the short normalization scheme development resource savings will be significant. But quality is medium.

All normalized fields and highlighted in green.



Good event normalization example



Bad event normalization example

In red highlighted fields that does not normalized because in parser missed some regex.

Normalization fields examples

- Username;
- Source IP(Attacker Address);
- Destination IP(Target Address);
- Protocol;
- Request URL;

- Asset Name.

Comparison

Feature	Qradar SIEM	ArcSight ESM/Express	Comments
Number of normalized fields	22	~400	
Add custom fields	Y	Y	ArcSight have limited number of reserved fields
Change built-in parsers	N	N	Only vendor can change built-in parsers
Extend built-in parsers	Y	Y	
Quality of normalization	Medium	Good	
Free access to built-in parsers	N	N	
Writing parser for your own application	Y	Y	