Forwarding log messages to Splunk from syslog-ng

How to configure syslog-ng to integrate with Splunk

syslog-ng
by ONE IDENTITY

Preface

Splunk is a popular search and analysis platform. Many users of Splunk also have syslog-ng™ deployed in their environments. This guideline describes some scenarios in which Splunk users can benefit from features of syslog-ng Premium Edition and the syslog-ng Store Box appliance and offers some technical guidance to optimize the syslog-ng™ configuration.

The following diagram presents a high-level overview of how using syslog-ng as your single, centralized log management solution can greatly enhance the benefits you achieve from Splunk or any other SIEM tool.

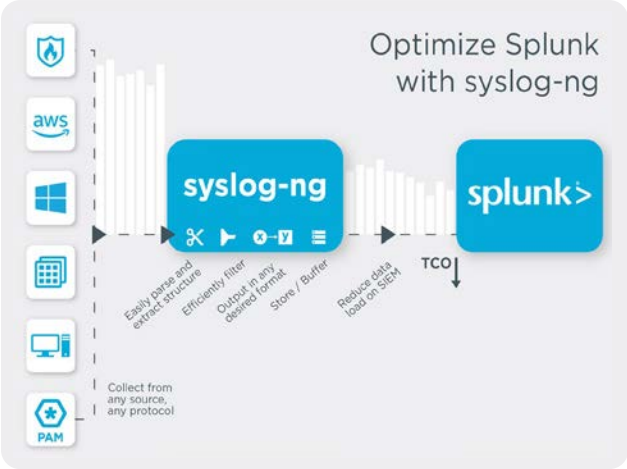

Figure 1. *Enhancing SIEM Efficiency and Results*

syslog-ng™ use cases for Splunk

This document describes the following scenarios and use cases.

Collecting logs from network devices

Major router manufacturers transfer log messages using the syslog protocol and syslog-ng™ natively supports both versions RFC3164 and RFC5424. Using syslog-ng™ can improve the reliability log data collection from network devices.

Long-term log storage

Organizations are required to archive data for compliance purposes, often for months or even years. Using syslog-ng™ reduces storage costs and secures log files.

Advanced filtering on clients to reduce data load

Many users use syslog-ng™ to filter log messages on clients to reduce network loads. Using syslog-ng™ can reduce the data load on Splunk, thereby improving performance and reducing license costs.

Routing to Splunk using the syslog-ng Store Box appliance

The syslog-ng Store Box provides log message collection, processing, storage and routing in an easy to set up and use appliance. Built upon the foundation of syslog-ng Premium Edition Version 7, the store box provides access to these capabilities via an intuitive browser-based graphical user interface.

Use case 1: Collecting logs from network devices

Collecting and centralizing log messages from network devices such as routers is one of the most common requirements addressed by syslog-ng™ with Splunk. Major router manufacturers like Cisco and Juniper use the syslog protocol to transfer event data. The syslog-ng™ application natively supports the original syslog protocol RFC3164 (also known as legacy-syslog or BSD-syslog) and the new syslog protocol RFC5424 (also known as IETF-syslog). In addition, syslog-ng™ also supports variants of these protocols that are used by certain router manufacturers.

Many routers only transport logs through the inherently unreliable UDP or are set to use UDP because of the perceived advantage of its lower resource requirements. However, this comes at a price. UDP is a best-effort datagram delivery service. UDP datagrams are not tracked or acknowledged; dropped datagrams will not be retransmitted. In contrast, TCP provides a much more robust network protocol. TCP implements a virtual circuit providing a byte stream service. It maintains segment sequence numbers, acknowledges received segments and retransmits missing ones. TCP inherently controls the flow of segments between peers and has control mechanisms to reduce or eliminate network congestion. Modern, optimized OS TCP/IP stacks combined in many cases with hardware

TCP/IP offload processing make TCP far more efficient than when the syslog protocol was initially developed (1980s).

Nevertheless, in most IT environments it is still necessary for a syslog server to accommodate high volumes of messages transmitted by UDP. Syslog-ng Premium Edition includes features to sustain very high UDP syslog datagram input volumes with minimal loss.

- Very high overall performance for log processing. Depending on the hardware used and the syslog-ng configuration, a single syslog-ng instance can typically ingest and process over half a million events per second (EPS).
- Relays. Relay instances of syslog-ng PE sited network-topologically close to the UDP sources can reliably ingest the UDP messages with low loss due to proximity (fewer router hops). The relays can then forward the messages to a syslog-ng PE server instance using TCP/IP. This will drastically increase reliability, especially if this leg will have a lot of router hops or traverses a wide-area network.
- Dedicated, high-performance UDP source available exclusively in syslog-ng PE. This source, named the udp-balancer(), leverages the extended Berkeley Packet Filter (eBPF) to evenly balance incoming syslog UDP datagrams into multiple, parallel backend sockets. The number of parallel threads thus available can be up to the number of virtual CPUs in the OS platform. This feature helps to eliminate a common cause of lost messages in which bursts of UDP datagrams fill the kernel's input buffer. When this happens, subsequent UDP datagrams arriving at the buffer are dropped. With the multiple, parallel, load-balanced back-end sockets made available by the udp-balancer() source, the kernel input buffer can be drained very quickly, minimizing or eliminating input buffer spills.

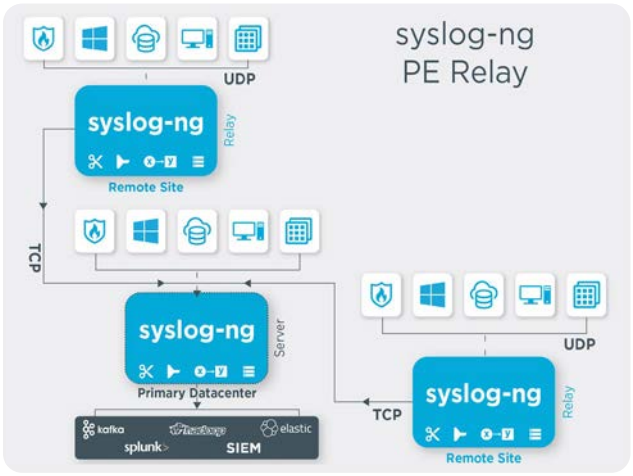

Figure 2. *Using Relays for Reliable UDP Ingestion*

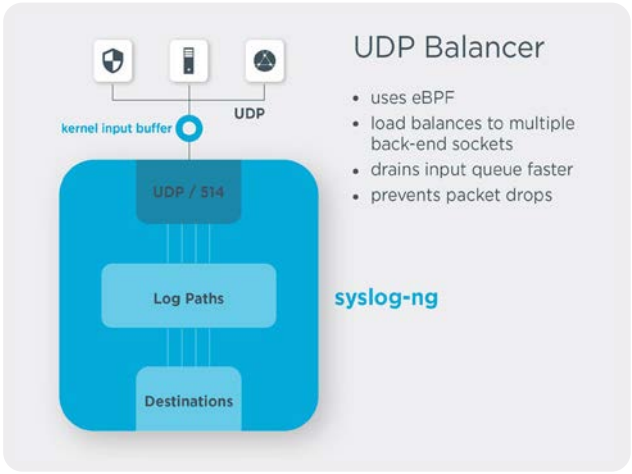

Figure 3. *High performance UDP ingestion with udp-balancer*

```
source s_udp_bsd-514_lb {
      udp-balancer(
        listeners(8)
        port(514)
        so-rcvbuf(16777216)
        log-fetch-limit(20000)
        log-iw-size(30000)

};
```

Figure 4. *syslog-ng PE udp-balancer source block*

Use case 2: Long-term log storage

Depending on the type of log messages being collected, organizations are required to archive data for compliance purposes. Many data retention policies and regulations specify that log messages be stored in their original format for several months or even years. If organizations do not need to analyze these data but simply must archive them securely, syslog-ng PE provides a cost-effective and convenient solution.

Users can specify the type of destination for archiving: text file, binary logstore, or SQL database. Output messages can be written to a specific file (or set of files) depending on certain criteria. The use of syslog-ng PE logstore facilitates both confidentiality and long-term archiving. It uses compression for saving space, while timestamping and encryption is used to ensure tamper-proof log storage.

By default, syslog-ng™ parses all incoming messages, creates name-value pairs and stores messages by reconstructing the message with the help of templates. For non-standard compliant messages, the stored messages might be slightly different from the original. With the store-raw-message flag on inputs, syslog-ng PE saves the original message as received from the client in a name-value pair called ${RAWMSG}. If you need the unmodified original message, you can use this macro both for long-term storage or forwarding to Splunk.

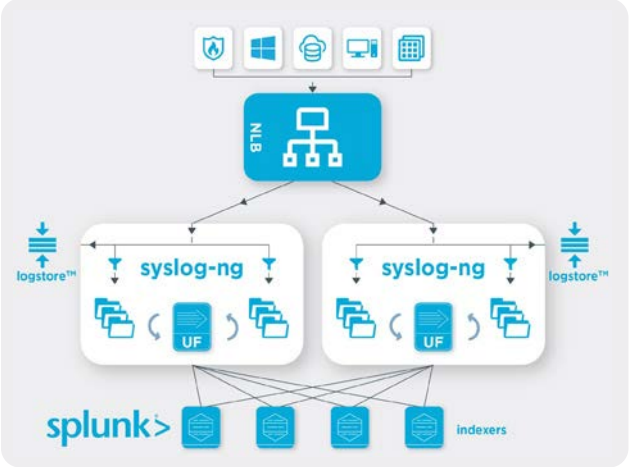

Figure 5. *Using logstore on syslog-ng for long-term storage*

```
destination d_logstore_enc {
  logstore( "/var/log/longterm/$MONTH$DAY.lgs"
            encrypt-certificate("/opt/syslog-ng/
etc/tls/server.crt")
            compress(3));
};
```

Figure 6. *PE logstore compressed, binary log storage destination*

The following configuration file example extends the previous example of receiving log messages from UDP by adding a long-term destination to the configuration called d_longterm. You can delete flat files regularly once they have been read by Splunk, while encrypted and compressed logstore files stay as long as required by compliance regulations. This file-based approach represents the previous method of using the Splunk Universal Forwarder (which is still supported by Splunk).

```
@version: 8.0
@include "scl.conf"
source s_net {
    udp-balancer(
     listeners(8)
     port(514));
};
destination d_files_splunk { file("/var/log/
splunk/$HOST/$MONTH$DAY.log" create_dirs(yes));
};
destination d_longterm { logstore("/var/log/
longterm/$MONTH$DAY.lgs"
    encrypt_certificate ("/opt/syslog-ng/etc/
syslog-ng/keys/public-server-certificate.pem"));
};
 log {
    source(s_net);
    destination(d_files_splunk);
    destination(d_longterm);
};
```

Figure 7. *Configuration of syslog-ng logstore for long term storage*

Figure 8 shows the current way of collecting logs for Splunk: Sending them with syslog-ng directly using Splunk HEC (Http Event Collector).

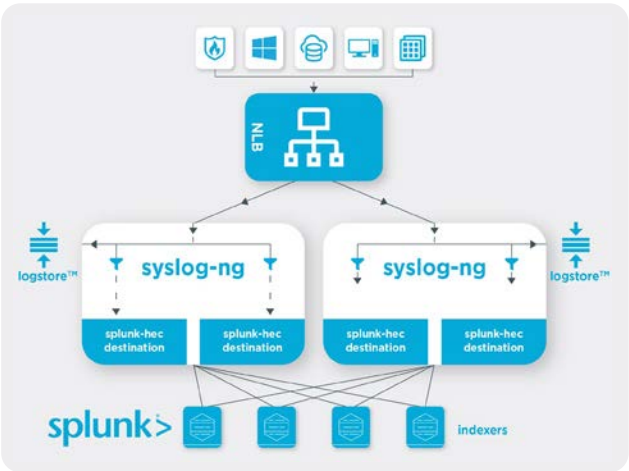

Figure 8. *Forwarding to Splunk using Splunk-HEC*

Use case 3: syslog-ng PE advanced filtering to reduce data load

Many users use syslog-ng™ to filter log messages on clients to reduce network loads. The syslog-ng™ application can filter out irrelevant content when network capacity to remote clients is limited. When defining a logpath, a user can insert a filter to route messages based on pre-defined criteria. Messages coming from the sources listed in the log statement and matching all the filters are sent to the listed destinations. To define a log path, add a log statement to the syslog-ng™ configuration file using the following syntax:

```
log {
      source(s1); source(s2); ...
      optional_
element(filter1|parser1|rewrite1);
      optional_
element(filter2|parser2|rewrite2);...
      destination(d1); destination(d2); ...
      flags(flag1[, flag2...]);
};
```

Figure 9. *Using filters (and other processing statements) in a log path*

The syslog-ng™ application can handle embedded log statements (also called log pipes). Embedded log statements are useful for creating complex, multi-level log paths with several destinations and for using filters, parsers and rewrite rules. For example, if you want to filter your incoming messages based on the facility parameter, then use further filters to send messages arriving from different hosts to different destinations, you would use embedded log statements. This advanced filtering enables users to fine-tune the number and type of messages sent to Splunk instances to be indexed (thereby lowering network capacity requirements).

Example: Filtering log messages

This example illustrates how to configure your syslog-ng™ if you wish to employ various filters to control the type and number of messages that get forwarded to Splunk.

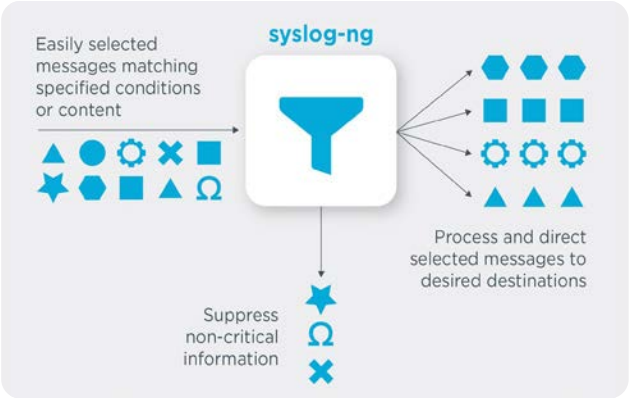

Figure 10. *syslog-ng filtering*

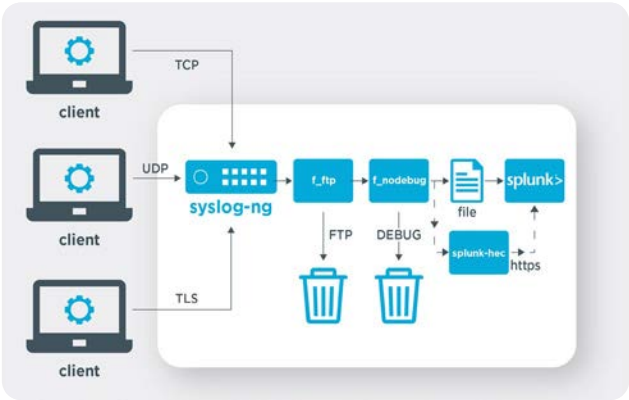

Figure 11. *Filtering in syslog-ng*

The following configuration file example extends the example of receiving logs via UDP by adding filters to the configuration. The first one is called f_proftpd and discards any messages from the application called proftpd. The second one Is called f_nodebug. This filter discards debug messages from the logs that are only necessary under very special circumstances, but can increase log volume considerably.

Another change to the configuration is that we use the splunk-hec() destination of syslog-ng PE here. As you can see, there are considerably fewer mandatory configuration options to specify compared with using the lower level http() destination directly. For a complete list of options, check the syslog-ng PE documentation. You can add options to scale and secure your logging infrastructure. For example, you can enable additional reliability using the disk-buffer, load balance data among multiple Splunk indexers, and more.

```
@version: 8.0
@include "scl.conf"
source s_net {
    udp-balancer(
    listeners(8)
    port(514));
};
destination d_splunk_hec {
     splunk-hec(
     index("main")
     token("fcddc233-a7c4-43fa-903a-
0654622c5093")
     url("http://your-splunk-server:8088/
services/collector/event") );
};
filter f_program { not program('proftpd');};
filter f_nodebug { level(info..emerg);};
log {
    source(s_net);
    filter(f_program);
    filter(f_nodebug);
    destination(d_splunk_hec);
};
```

Figure 12. *syslog-ng PE configuration using splunk-hec destination*

Use case 4: syslog-ng Store Box

As previously described, the syslog-ng Store Box (SSB) appliance is built on syslog-ng Premium Edition (PE). The SSB inherits most of syslog-ng PE's features and makes them available with an easy-to-use graphical user interface. One of the typical use cases for the SSB, just as for PE, is optimizing the logging infrastructure for SIEM / log analysis tools. The SSB has a built-in destination for Splunk which uses Splunk HEC as the communications protocol. As with other standard SSB destinations, this is easily configured via a simple to use dialog presented through its intuitive web UI.

Like its PE sibling, the syslog-ng Store Box appliance can collect log messages from many different log sources, in many formats. These include UNIX / Linux / Windows system logs, firewall and router logs, various application logs, and now SQL sources as well. SSB can parse, rewrite, filter, and store log messages. In addition to the traditional syslog-ng features, the SSB appliance provides an interface to search log messages, and does complete log life cycle management, including archiving and backup. Finally, it can also forward events to various on-prem and cloud destinations. It allows you to optimize your SIEM installations both for resources and licensing, as you can collect log messages in a single step, store them on SSB, and only forward a reduced subset of logs to various analytics tools.

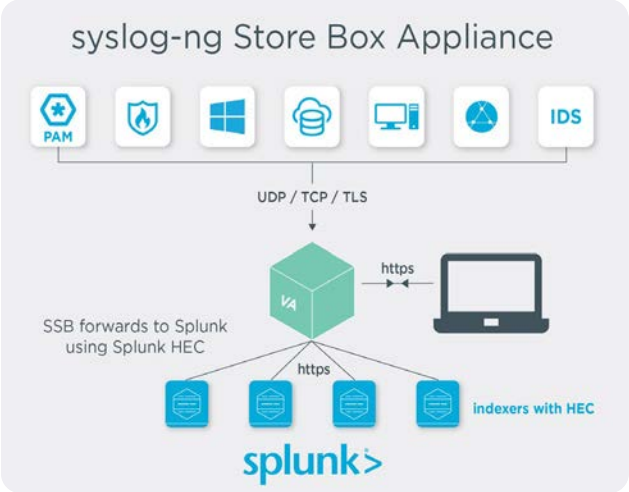

Figure 13. *SSB forwarding to Splunk using Splunk HEC*

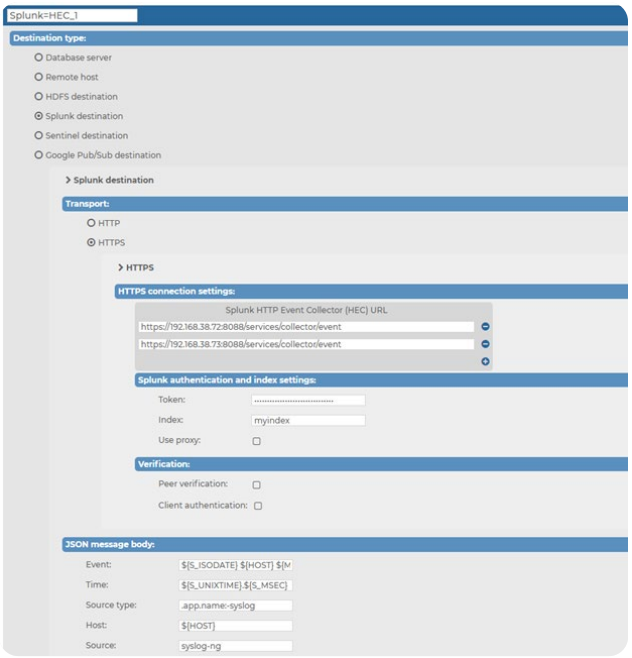

Figure 14. *Splunk-HEC Configuration using SSB GUI*