

# Table of Contents

<b>Сборка IPFIX потоков rcollector2 .....</b>	<b>3</b>
<b><i>Введение .....</i></b>	<b>3</b>
<b><i>Установка и обновление .....</i></b>	<b>3</b>
<b><i>Файлы поставки .....</i></b>	<b>3</b>
<b><i>Параметры запуска программы .....</i></b>	<b>4</b>
<b><i>Конфигурация .....</i></b>	<b>4</b>
<b><i>Статистика работы программы .....</i></b>	<b>8</b>



# Сборка IPFIX потоков rcollector2

## Введение

Утилита предназначена для дополнения данными вспомогательных потоков таких как clickstream, SIP из потока сессий (netflow). По умолчанию полученные данные сохраняются в файлы, но данные так же можно сохранять и в БД ИС СОРМ-3.

## Установка и обновление

1. подключите репозиторий VAS Experts аналогично п.1 инструкции [установки DPI](#).
2. установите rcollector2:

```
yum install -y rcollector2
```

3. настройте файлы конфигурации в директории /etc/rcollector2 см. далее



Внимание! При переходе с предыдущей версии [rcollector](#) на rcollector2 необходимо внести изменения в скрипты запуска.

## Файлы поставки

1. примеры конфигурации:

```
/etc/rcollector2/rc2flowprocess - пример исполняемого файла для обработки сессий (netflow)
/etc/rcollector2/rc2urlprocess - пример исполняемого файла для обработки HTTP сессий
/etc/rcollector2/rc2sipprocess - пример исполняемого файла для обработки SIP сессий
/etc/rcollector2/rcollector_flow.properties - пример конфигурационного файла для режима работы flow
/etc/rcollector2/rcollector_url.properties - пример конфигурационного файла для режима работы url
/etc/rcollector2/rcollector_sip.properties - пример конфигурационного файла для режима работы sip
```

2. исполняемый файл:

```
/bin/rcollector2
```

# Параметры запуска программы

Программа rcollector2 имеет следующие параметры запуска:

```
usage: rcollector2 ОПЦИИ
```

ОПЦИИ:

- -h, --help отобразить краткую справку.
- -fCONFIG, --config-file=CONFIG конфигурационный файл.
- -mMODE, --mode=MODE режим работы программы. Режимы: flow, urlget, sipget
- -uidUNIQUEID, --uniqueid=UNIQUEID уникальный номер точки обработки.
- -ifINFILE, --infile=INFILE входной файл.
- -ofOUTFILE, --outfile=OUTFILE выходной файл.
- -asnASN, --localasn=ASN список локальных автономных систем (разделенных запятой)
- -oufOUTFILTER, --outfilter=OUTFILTER вывод отдельных записей по dpi protocol id в файл. Формат: <outfile>,<protocol>,...<protocol>|[next filter]. Пример: telnet.dump,22,23
- -tdUSEFILTER, --tordb=USEFILTER путь к файлу с IP адресами сети TOR. По умолчанию /var/data/tor/ip.db.gz.
- -sdbDIR, --sessiondb=DIR путь к каталогу с данными о сессиях. По умолчанию /var/db/rcollector.
- -outmailFILE, --outfilemail=FILE выходной файл для данных о mail соединениях в режиме sipget.
- -outftpFILE, --outfileftp=FILE выходной файл для данных о ftp соединениях в режиме sipget.
- -outimFILE, --outfileim=FILE выходной файл для данных о im соединениях в режиме sipget.
- -dhINTEGER, --depth=INTEGER глубина поиска фалов с данными о сессиях.
- -sdr, --session-db-read-thread использовать несколько потоков при загрузке данных о сессиях.



Внимание! В некоторых случаях, например, при частом сбросе кеша ОС, данная опция может существенно увеличить время загрузки данных о сессиях. При использовании данной опции, по умолчанию, создаются 2 потока для чтения данных. **В примерах исполняемых файлов указана данная опция.**

- -v, --version вывести версию программы.

## Конфигурация

Параметры работы программы задаются в .properties файле. По умолчанию загружается конфигурационный файл из каталога /etc/rcollector2 с именем rcollector\_MODE.properties, где MODE это выбранный режим работы. Для режима flow - flow; для режима urlget - url; для режима sipget - sip.



При вставке данных в БД, в конфигурационном файле **обязательно** должны быть заданы следующие параметры:

- db.host

- db.port
- db.user
- db.pass
- db.name
- db.telco\_code
- db.bad\_rows\_dir
- db.validation\_error\_path

При вставке данных в БД выходные файлы не создаются. В случае отсутствия подключения к БД будут создаваться выходные файлы согласно параметрам командной строки. В процессе работы может создаваться файл в каталоге из параметра db.validation\_error\_path, который содержит краткую информацию об отброшенных данных, не прошедших проверку на наличие необходимых полей. Имя файла соответствует имени входного файла с добавлением расширения .err .

### Параметр cachedb

Данный параметр позволяет настроить работу с файлами данных о сессиях.

- max\_reader\_threads - максимальное количество потоков, запускаемых одновременно, для чтения данных о сессиях из файлов. Целое число от 0 до 6.



В некоторых случаях слишком большое количество потоков может привести к замедлению загрузки файлов, что приведет к общему замедлению обработки данных.

### Параметр stats

Данный параметр устанавливает возможность отправки статистики работы программы в telegraf.

- stats.socket\_path - путь к datagram socket telegraf'a.
- stat.stag - тег, выставляемый в поле rcollector\_tag при отправке статистики в telegraf.

### Параметр db

Данный параметр позволяет организовать вывод полученных данных в БД ИС СОРМ-3.

- db.host - адрес сервера postgresql
- db.port - порт
- db.user - имя пользователя
- db.pass - пароль пользователя
- db.name - имя БД
- db.bad\_rows\_dir - каталог для размещения файлов с данными в формате PGCOPY, которые были отвергнуты сервером postgresql
- db.validation\_error\_path - каталог для файлов с описанием причины для отброшенных входных данных
- db.copy\_threads - количество потоков, которые вставляют данные в postgresql используя

COPY в бинарном формате, по умолчанию 1

- db.commit\_rows - количество строк в одном блоке, отправляемом на запись в бд при использовании COPY, по умолчанию 5000
- db.telco\_code - идентификатор telco для записи в соответствующее поле бд
- db.llds\_id - идентификатор типа источника, по умолчанию устанавливаются следующие значения:
  - для режима flow - 309
  - для режима urlget - 310
  - для режима sipget - 311
- db.ftp.llds\_ldst\_id - идентификатор типа источника для режимов flow и sipget при вставке ftp данных, по умолчанию 307
- db.email.llds\_ldst\_id - идентификатор типа источника для режима sipget при вставке email данных, по умолчанию 304
- db.im.llds\_ldst\_id - идентификатор типа источника для режима sipget при вставке im данных, по умолчанию 306
- db.terminal.llds\_ldst\_id - идентификатор типа источника для режима flow при вставке terminal данных, по умолчанию 308
- db.h323.llds\_ldst\_id - идентификатор типа источника для режима flow при вставке h323 данных, по умолчанию 311
- db.ftp\_proto - идентификаторы для определения данных как ftp и их занесение в БД, по умолчанию "20,21,69,115,152,349,574,662,989,990,3713,6620,6621,6622,65086". Для отключения необходимо указать none.
- db.ssh\_proto - идентификаторы для определения данных как terminal и их занесение в БД, по умолчанию "22,23,3820,992,220". Для отключения необходимо указать none.
- db.h323\_proto - идентификаторы для определения данных как h323 и их занесение в БД, по умолчанию "4569,49217". Для отключения необходимо указать none.
- db.require\_subscriber\_id - проверять наличие subscriber\_id во входных данных, по умолчанию true. Если subscriber\_id будет отсутствовать и параметр выставлен в true, то запись будет отброшена, о чем будет сообщено в информационном файле
- db.http.length.htrq\_url - максимальное количество символов для поля htrq\_url. По умолчанию 1024
- db.ftp.length.ftpc\_server\_name - максимальное количество символов для поля ftpc\_server\_name. По умолчанию 256
- db.ftp.length.ftpc\_user\_name - максимальное количество символов для поля ftpc\_user\_name. По умолчанию 64
- db.ftp.length.ftpc\_user\_password - максимальное количество символов для поля ftpc\_user\_password. По умолчанию 256
- db.email.length.emlc\_sender - максимальное количество символов для поля emlc\_sender. По умолчанию 256
- db.email.length.emlc\_subject - максимальное количество символов для поля emlc\_subject. По умолчанию 256
- db.email.length.emlc\_reply\_to - максимальное количество символов для поля emlc\_reply\_to. По умолчанию 256
- db.email.length.emcr\_receiver - максимальное количество символов для поля emcr\_receiver. По умолчанию 256
- db.email.length.mlcs\_server - максимальное количество символов для поля mlcs\_server. По умолчанию 256
- db.im.length.imcn\_user\_login - максимальное количество символов для поля imcn\_user\_login. По умолчанию 20
- db.im.length.imcn\_user\_password - максимальное количество символов для поля imcn\_user\_password. По умолчанию 16

- db.im.length.imcn\_sender\_screen\_name - максимальное количество символов для поля imcn\_sender\_screen\_name. По умолчанию 32
- db.im.length.imcn\_sender\_uin - максимальное количество символов для поля imcn\_sender\_uin. По умолчанию 256
- db.im.length.imcr\_receiver\_screen\_name - максимальное количество символов для поля imcr\_receiver\_screen\_name. По умолчанию 32
- db.voip.length.vipc\_conference\_id - максимальное количество символов для поля vipc\_conference\_id. По умолчанию 64
- db.voip.length.vipc\_originator\_name - максимальное количество символов для поля vipc\_originator\_name. По умолчанию 64
- db.voip.length.vipc\_calling\_original\_number - максимальное количество символов для поля vipc\_calling\_original\_number. По умолчанию 128
- db.voip.length.vipc\_called\_original\_number - максимальное количество символов для поля vipc\_called\_original\_number. По умолчанию 128
- db.do\_content\_id - флаг, позволяющий сохранять dpi session\_id в полях data\_content\_id бд. По умолчанию false

## Параметр nat

Данные параметры позволяют дополнить flow данными о трансляциях адресов в случае их отсутствия во входном файле.

- nat.sessions\_dir - каталог для поиска файлов трансляций NAT. Для обработки берутся последние по времени создания файлы. Маска поиска файлов url\_\*.dump, url\_\*.dump.gz.
- nat.files\_cnt - количество файлов, которые будут использованы для обработки. По умолчанию 1 файл.

Файл трансляций должен быть в формате csv с символом разделения табуляция и иметь следующий формат полей:

№ поля	Описание
1	Время трансляции
2	Протокол
3	Тип события NAT
4	IP адрес источника
5	Порт источника
6	IP адрес источника после NAT
7	Порт источника после NAT

## Параметр logging

Данный параметр отвечает за настройку логирования программы.

- logging.loggers.root.level - уровень логирования
- logging.loggers.root.channel - канал для вывода сообщений
- logging.channels.fileChannel.class - класс канала вывода
- logging.channels.fileChannel.path - путь к лог-файлу
- logging.channels.fileChannel.rotation - параметр ротации
- logging.channels.fileChannel.archive - параметр имени архивных файлов

- logging.channels.fileChannel.purgeCount - количество архивных файлов
- logging.channels.fileChannel.formatter.class - класс форматировщика
- logging.channels.fileChannel.formatter.pattern - шаблон для форматировщика
- logging.channels.fileChannel.formatter.times - время



Более подробно ознакомиться с параметрами логирования можно по ссылке [Class FileChannel](#).

## Статистика работы программы

Типы полей статистических данных о работе программы.

### Режим sip

- read\_lines - количество прочитанных строк входного файла
- sip\_bye - количество записей SIP BYE
- sip\_invite - количество записей SIP INVITE
- sip\_miss - количество записей, не имеющих информацию в кэше соединений
- count\_ftp - количество записей ftp
- bad\_ftp - количество ftp записей не сохраненных в файл
- out\_ftp - количество ftp записей успешно сохраненных в файл
- dup\_ftp - количество дублированных ftp записей
- count\_mail - количество записей mail
- bad\_mail - количество mail записей не сохраненных в файл
- out\_mail - количество mail записей успешно сохраненных в файл
- dup\_mail - количество дублированных mail записей
- count\_im - количество записей im
- bad\_im - количество im записей не сохраненных в файл
- out\_im - количество im записей успешно сохраненных в файл
- bad\_sip - количество sip записей не сохраненных в файл
- out\_sip - количество sip записей успешно сохраненных в файл
- dup\_sip - количество дублированных sip записей
- work\_time - время работы программы в миллисекундах

### Режим url

- read\_lines - количество прочитанных строк входного файла
- sess\_miss - количество записей для которых нет информации в данных о сессиях
- resp\_miss - количество записей для которых нет информации в данных об ответах
- resp\_skip - количество отброшенных записей (эти записи ответы от серверов)
- out\_lines - количество сохраненных строк в выходном файле
- work\_time - время работы программы в миллисекундах

### Режим flow



- `read_lines` - количество прочитанных строк входного файла
- `marked_as_tor` - количество записей, промаркированных как TOR
- `out_lines` - количество сохраненных строк в выходном файле
- `work_time` - время работы программы в миллисекундах