Description of HTTP Output File Formats

There are two basic file formats that are used for the output of the HTTP trace analysis tools; one that is sorted by TCP connection (output `of http_connect`) and one that is sorted by client host (output of `http_active`). Both types are text files with very simple fixed-field format where fields are separated by one or more spaces. Each line in a file represents summary information about some "event" that has been extracted from the raw `tcpdump` traces. Timestamps in these files have the format ssssss.mmmmmm and are interpreted as fractional seconds with microsecond resolution.

`http_connect` Output Format

The fields in this file in left to right order:

- Timestamp#1 (the meaning is conditional on the value of the Event field as described below)
- Client (browser) IP address
- Client (browser) Port number
- ">"
- Server IP address
- Server Port number":" (the colon is there only for use by some of our tools)
- Event (Event may have one of the following values)

    - "SYN" – the normal beginning of a TCP connection; timestamp#1 is the connection start time (timestamp of the SYN packet); timestamp#2 is blank
    - "ACT-REQ" – the beginning of a TCP connection where the tcpdump did not contain SYN but the connection was found to have a request in progress; timestamp#1 is the timestamp of the first packet found in the request; timestamp#2 is blank
    - "ACT-RSP" – the beginning of a TCP connection where the tcpdump did not contain SYN but the connection was found to have a response in progress; timestamp#1 is the timestamp of the first packet found in the response; timestamp#2 is blank
    - "FIN" – the normal end of a TCP connection with a FIN; timestamp#1 is the timestamp of the last packet seen in the connection; timestamp#2 is the connection logical end time (timestamp of the FIN packet)
    - "RST" – the normal end of a TCP connection with a Reset; timestamp#1 the timestamp of the last packet seen in the connection; timestamp#2 is the connection logical end time (timestamp of the Reset packet)
    - "TRM" – the end of a TCP connection without normal completion (tcpdump did not have a FIN or Reset); timestamp#1 is the timestamp of the last packet seen in the connection; timestamp#2 is the same as timestamp#1
    - "REQ" – a browser request; timestamp#1 is the request *start* time (timestamp of first packet containing part of the request); timestamp#2 is the request end time (timestamp of last packet containing part of the request.
    - "RSP" – a server response; timestamp#1 is the response *end* time (time of last packet containing part of the request); timestamp#2 is the response start time (timestamp of first packet containing part of the request.
    - "ERR:" – an exceptional condition was detected during analysis of the connection; timestamp#1 is the time of the packet where the condition was detected; the last two fields in the line are replaced by a text description of the condition.

- Request or Response size in number of bytes (or blank for other non-error event types)
- Timestamp#2 (the meaning is conditional on the value of the Event field as described above)

NOTE: A sequence of lines with the same values for Client IP address/port and Server IP address/port are all for multiple events from the same TCP connection.

`http_active` Output Format

The fields in this files in left to right order:

- Timestamp#1 (the meaning is conditional on the value of the Event field as described below)
- Client (browser) IP address
- Client (browser) Port number (or "*" if no port is associated with the event)
- ">"
- Server IP address (or "*" if no server is associated with the event)
- Server Port number":" (or "*" if no server is associated with the event) (the colon is there only for use by some of our tools)
- Event (Event may have one of the following values)

    - "REQ" – a browser request
        - timestamp#1 is the request *start* time (timestamp of first packet containing part of the request)
        - timestamp#2 is the request end time (timestamp of last packet containing part of the request)
        - event-data is the request size in number of bytes

    - "RSP" – a server response
        - timestamp#1 is the response *end* time (time of last packet containing part of the request)
        - timestamp#2 is the response start time (timestamp of first packet containing part of the request)
        - event-data is the response size in number of bytes

    - "IDLE" – a period of time longer than a threshold value (default 2 seconds) no activity by the browsing client because either (a) there were no TCP connections or (b) no existing TCP connection had a request/response exchange in progress
        - timestamp#1 is ending time of the idle period
        - timestamp#2 is the beginning time of the idle period
        - event-data is the elapsed time of the idle period in milliseconds.

    - "ERR:" (or "ERROR") – an exceptional condition was detected during analysis of the connection; timestamp#1 is the time of the packet where the condition was detected; the last two fields in the line are replaced by a text description of the condition.

- Event-Data (the meaning is conditional on the value of the Event field as described above)
- Timestamp#2 (the meaning is conditional on the value of the Event field as described above)

NOTE: A sequence of lines with the same value for Client IP Address are all for multiple events from the same browsing client.