Improve smbcmp the capture diff tool

Google Summer of Code 2019

Mairo P. Rufus <akoudanilo@gmail.com>
Mentor: Aurélien Aptel <aaptel@suse.com>

Who am I

- Master in Computer Science student
- at Polytechnic Yaounde, Cameroon
- Graduating this year
- github.com/rmpr
- @rmpr@hostux.social

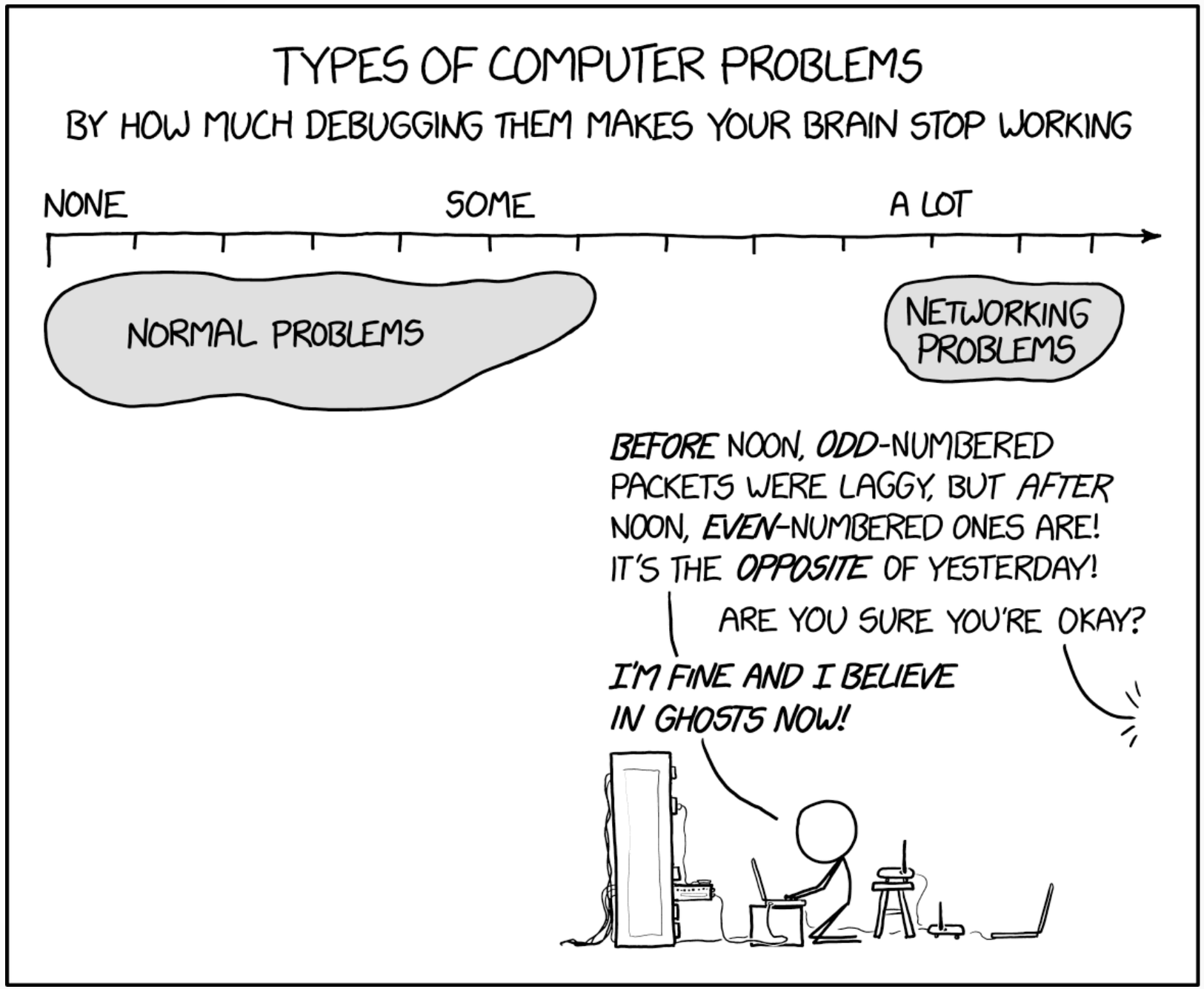
Useful Links

- Repository: github.com/smbcmp/smbcmp
- SambaXP 2018: sambaxp.org/fileadmin/user_upload/sambaXP2018-Slides/aaptel-smbcmp.pdf
- SDC 2019: youtube.com/watch?v=H4z-2iHVuwg
- LCA 2020: youtube.com/watch?v=6yhKWq3-sr4

Content

- What is the GSOC?
- What is smbcmp?
- Choosing the PDML output of Tshark
- GUI for smbcmp
- Port to other platforms

Networking problems are hard to debug… xkcd 2259



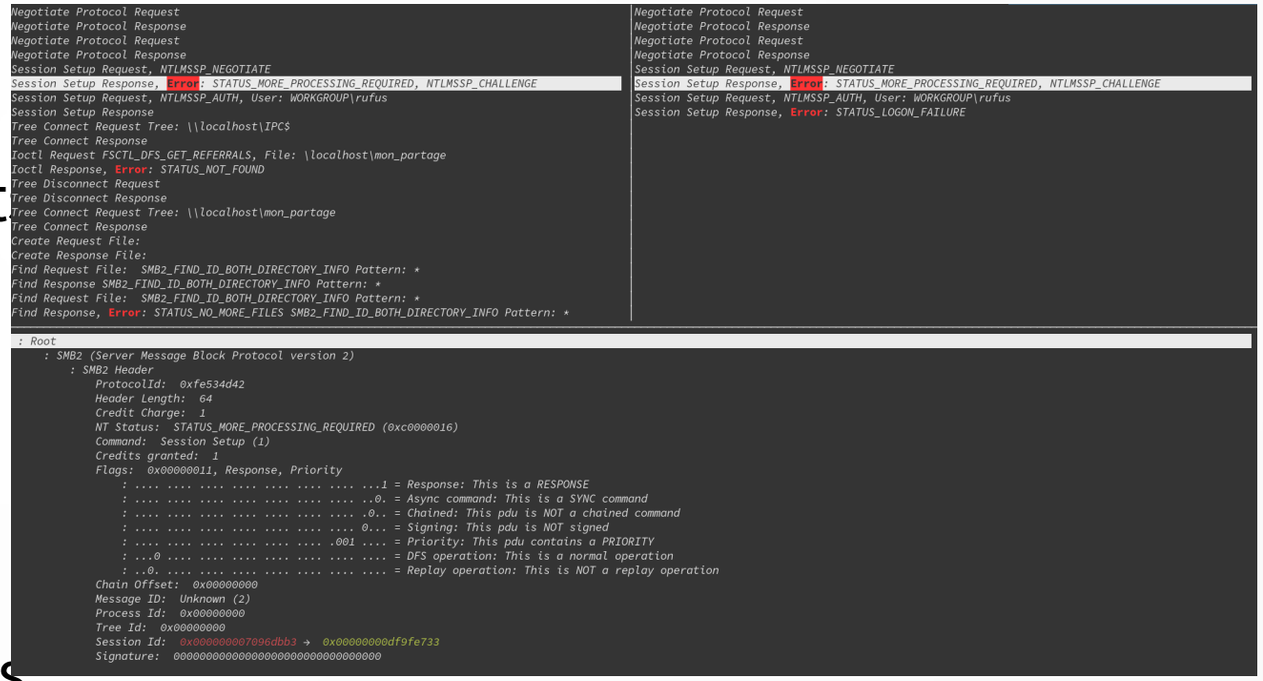
What is the GSOC?

- Global program for 18+ years old students
- Each student works on an OSS project for an org
- Each student is assigned at least one mentor
- The programs lasts for 3 months

find more at : summerofcode.withgoogle.com

What is smbcmp?

- Network capture diff for SMB
- Supports Encrypted SMB packet
- Uses Tshark in the background
- 2 modes: Single Trace, Diff traces



Tshark's text output (-V)

```
SMB2 (Server Message Block Protocol version 2)
    SMB2 Header
        ProtocolId: 0xfe534d42
        Header Length: 64
        Credit Charge: 0
        NT Status: STATUS_SUCCESS (0x00000000)
        Command: Negotiate Protocol (0)
        Credits granted: 1
        Flags: 0x00000001, Response
            .... .... .... .... .... .... .... ...1 = Response: This is a RESPONSE
            .... .... .... .... .... .... .... ..0. = Async command: This is a SYNC command
            .... .... .... .... .... .... .... .0.. = Chained: This pdu is NOT a chained command
            .... .... .... .... .... .... .... 0... = Signing: This pdu is NOT signed
            .... .... .... .... .... .... .000 .... = Priority: This pdu does NOT contain a PRIORITY
            ...0 .... .... .... .... .... .... .... = DFS operation: This is a normal operation
            ..0. .... .... .... .... .... .... .... = Replay operation: This is NOT a replay operation
        Chain Offset: 0x00000000
        Message ID: Unknown (0)
        Process Id: 0x00000000
        Tree Id: 0x00000000
        Session Id: 0x0000000000000000
        Signature: 00000000000000000000000000000000
    Negotiate Protocol Response (0x00)
        StructureSize: 0x0041
            0000 0000 0100 000. = Fixed Part Length: 32
            .... .... .... ...1 = Dynamic Part: True
        Security mode: 0x01, Signing enabled
            .... ...1 = Signing enabled: True
            .... ..0. = Signing required: False
        Dialect: SMB2 wildcard (0x02ff)
        NegotiateContextCount: 0
        Server Guid: 61636f6c-686c-736f-7400-000000000000
        Capabilities: 0x00000007, DFS, LEASING, LARGE MTU
            .... .... .... .... .... .... .... ...1 = DFS: This host supports DFS
            .... .... .... .... .... .... .... ..1. = LEASING: This host supports LEASING
```

Tshark's PDML (-T pdml)

```
<proto name="smb2" showname="SMB2 (Server Message Block Protocol version 2)" size="202" pos="90">
  <field name="" show="SMB2 Header" size="64" pos="90" value="fe534d4240000000000000000000010001000000000000000000000000000000000000000000000000000000000000000000000000
    <field name="smb2.protocol_id" showname="ProtocolId: 0xfe534d42" size="4" pos="90" show="0xfe534d42" value="fe534d42"/>
    <field name="smb2.header_len" showname="Header Length: 64" size="2" pos="94" show="64" value="4000"/>
    <field name="smb2.credit.charge" showname="Credit Charge: 0" size="2" pos="96" show="0" value="0000"/>
    <field name="smb2.nt_status" showname="NT Status: STATUS_SUCCESS (0x00000000)" size="4" pos="98" show="0" value="00000000"/>
    <field name="smb2.cmd" showname="Command: Negotiate Protocol (0)" size="2" pos="102" show="0" value="0000"/>
    <field name="smb2.credits.granted" showname="Credits granted: 1" size="2" pos="104" show="1" value="0100"/>
    <field name="smb2.flags" showname="Flags: 0x00000001, Response" size="4" pos="106" show="0x00000001" value="01000000">
      <field name="smb2.flags.response" showname=".... .... .... .... .... .... .... ...1 = Response: This is a RESPONSE" size="4" pos="106" show="1" val
      <field name="smb2.flags.async" showname=".... .... .... .... .... .... .... ..0. = Async command: This is a SYNC command" size="4" pos="106" show="
      <field name="smb2.flags.chained" showname=".... .... .... .... .... .... .... .0.. = Chained: This pdu is NOT a chained command" size="4" pos="106"
      <field name="smb2.flags.signature" showname=".... .... .... .... .... .... .... 0... = Signing: This pdu is NOT signed" size="4" pos="106" show="0"
      <field name="smb2.flags.priority_mask" showname="... .... .... .... .... .... .... .000 .... = Priority: This pdu does NOT contain a PRIORITY" size="4"
      <field name="smb2.flags.dfs" showname="...0 .... .... .... .... .... .... .... = DFS operation: This is a normal operation" size="4" pos="106" show
      <field name="smb2.flags.replay" showname="..0. .... .... .... .... .... .... .... = Replay operation: This is NOT a replay operation" size="4" pos=
    </field>
    <field name="smb2.chain_offset" showname="Chain Offset: 0x00000000" size="4" pos="110" show="0x00000000" value="00000000"/>
    <field name="smb2.msg_id" showname="Message ID: Unknown (0)" size="8" pos="114" show="0" value="0000000000000000"/>
    <field name="smb2.pid" showname="Process Id: 0x00000000" size="4" pos="122" show="0x00000000" value="00000000"/>
    <field name="smb2.tid" showname="Tree Id: 0x00000000" size="4" pos="126" show="0x00000000" value="00000000"/>
    <field name="smb2.sesid" showname="Session Id: 0x0000000000000000" size="8" pos="130" show="0x0000000000000000" value="0000000000000000"/>
    <field name="smb2.signature" showname="Signature: 00000000000000000000000000000000" size="16" pos="138" show="00:00:00:00:00:00:00:00:00:00:00:00:00:
  </field>
  <field name="" show="Negotiate Protocol Response (0x00)" size="138" pos="154" value="41000100ff0200006c6f63616c686f737400000000000000070000000000800000
    <field name="smb2.buffer_code" showname="StructureSize: 0x0041" size="2" pos="154" show="0x00000041" value="4100">
      <field name="smb2.buffer_code.length" showname="0000 0000 0100 000. = Fixed Part Length: 32" size="2" pos="154" show="32" value="20" unmaskedvalue=
      <field name="smb2.buffer_code.dynamic" showname=".... .... .... ...1 = Dynamic Part: True" size="2" pos="154" show="1" value="1" unmaskedvalue="410
    </field>
    <field name="smb2.sec_mode" showname="Security mode: 0x01, Signing enabled" size="1" pos="156" show="0x00000001" value="01">
      <field name="smb2.sec_mode.sign_enabled" showname=".... ...1 = Signing enabled: True" size="1" pos="156" show="1" value="1" unmaskedvalue="01"/>
      <field name="smb2.sec_mode.sign_required" showname=".... ..0. = Signing required: False" size="1" pos="156" show="0" value="0" unmaskedvalue="01"/>
    </field>
    <field name="smb2.dialect" showname="Dialect: SMB2 wildcard (0x02ff)" size="2" pos="158" show="0x000002ff" value="ff02"/>
    <field name="smb2.negotiate_context.count" showname="NegotiateContextCount: 0" size="2" pos="160" show="0" value="0000"/>
    <field name="smb2.server_guid" showname="Server Guid: 61636f6c-686c-736f-7400-000000000000" size="16" pos="162" show="61636f6c-686c-736f-7400-0000000
    <field name="smb2.capabilities" showname="Capabilities: 0x00000007, DFS, LEASING, LARGE MTU" size="4" pos="178" show="0x00000007" value="07000000">
      <field name="smb2.capabilities.dfs" showname=".... .... .... .... .... .... .... ...1 = DFS: This host supports DFS" size="4" pos="178" show="1" va
      <field name="smb2.capabilities.leasing" showname=".... .... .... .... .... .... .... ..1. = LEASING: This host supports LEASING" size="4" pos="178"
      <field name="smb2.capabilities.large_mtu" showname=".... .... .... .... .... .... .... .1.. = LARGE MTU: This host supports LARGE MTU" size="4" pos
```

Tshark's Json (-T json)


```
        "smb2": {
          "SMB2 Header": {
            "smb2.protocol_id": "0xfe534d42",
            "smb2.header_len": "64",
            "smb2.credit.charge": "0",
            "smb2.nt_status": "0",
            "smb2.cmd": "0",
            "smb2.credits.granted": "1",
            "smb2.flags": "0x00000001",
            "smb2.flags_tree": {
              "smb2.flags.response": "1",
              "smb2.flags.async": "0",
              "smb2.flags.chained": "0",
              "smb2.flags.signature": "0",
              "smb2.flags.priority_mask": "0",
              "smb2.flags.dfs": "0",
              "smb2.flags.replay": "0"
            },
            "smb2.chain_offset": "0x00000000",
            "smb2.msg_id": "0",
            "smb2.pid": "0x00000000",
            "smb2.tid": "0x00000000",
            "smb2.sesid": "0x0000000000000000",
            "smb2.signature": "00:00:00:00:00:00:00:00:00:00:00:00:00:00:00:00"
          },
          "Negotiate Protocol Response (0x00)": {
            "smb2.buffer_code": "0x00000041",
            "smb2.buffer_code_tree": {
              "smb2.buffer_code.length": "32",
              "smb2.buffer_code.dynamic": "1"
            },
            "smb2.sec_mode": "0x00000001",
            "smb2.sec_mode_tree": {
              "smb2.sec_mode.sign_enabled": "1",
              "smb2.sec_mode.sign_required": "0"
            },
            "smb2.dialect": "0x000002ff",
            "smb2.negotiate_context.count": "0",
            "smb2.server_guid": "61636f6c-686c-736f-7400-000000000000",
            "smb2.capabilities": "0x00000007",
            "smb2.capabilities_tree": {
              "smb2.capabilities.dfs": "1",
              "smb2.capabilities.leasing": "1",
              "smb2.capabilities.large_mtu": "1",
```

Why use another output?

- Make better, more precise diffs
  - Add ignore rules: hide field if field < value
  - More complicated rules: if field X > field Y highlight difference
- More detailed output

Tshark's formats pros/cons

Format	Pros	Cons
PDML	• XML based • C implementation of the library • Human readable field name (showname attribute)	• Irrelevant information (pos, size)
Json	• No irrelevant information • Easier to parse (Python's built-in dict)	• No summary lines • No human readable field name and description (e.g. "smb2.negotiate_context.hash_algorithm": "0x00000001") • JSON dictionnary entries are not ordered (< Python 3.6)

First try: xmldiff

github.com/Shoobx/xmldiff

- A library and command line utility for diffing xml
- Based on “Change Detection in Hierarchically Structured Information”: ilpubs.stanford.edu:8090/115/1/1995-46.pdf

First try: xmldiff

- Offers an API to use xmldiff as a Python library
- Possibility to choose many parameters:
  - Ratio mode: How accurately the similarities are computed
  - Fast match: Find chains of matching nodes
  - Formatter: Presentation of results

First try: xmldiff

- Difficulties
  - Without fast match → too slow
  - With fast match → not really accurate
  - Too much noise (comparison of packets not really related)
  - Pdml structure not suited to xmldiff (field names are attributes instead of tags)

  → Not reliable to compute pdml diffs on the fly

Solution:

- Come up with our own implementation (DFS):
  - Take advantage of the structure of a SMB packet
  - A simple heuristic: the "Command" field of the SMB header
  - When stumbling on a non-flat node, reuse difflib
  - Possibility to expand it with ignore rules

  SMB2 specification: winprotocoldoc.blob.core.windows.net/productionwindowsarchives/MS-SMB2/%5BMS-SMB2%5D.pdf

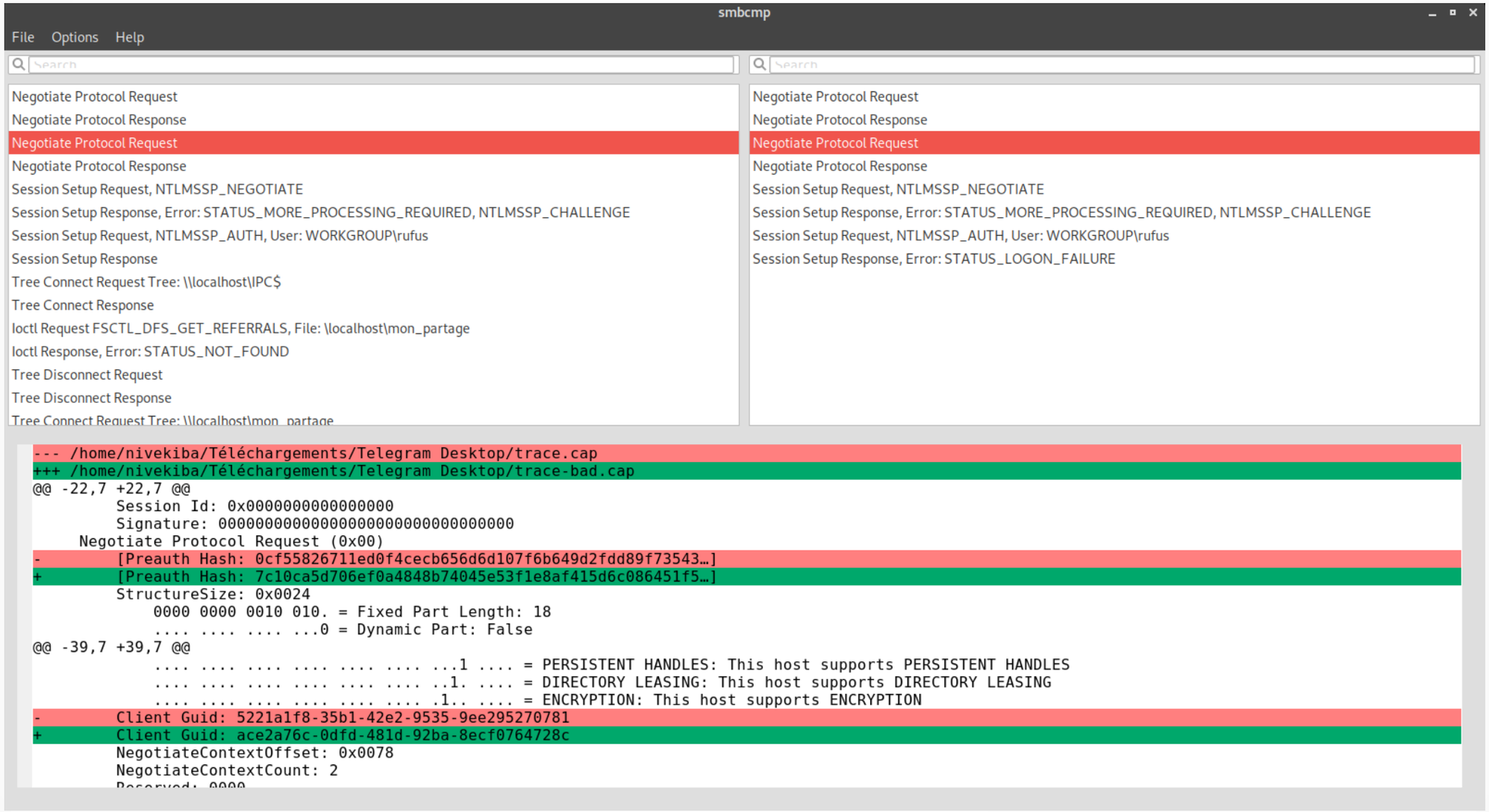
Why a GUI?

- More control on diff presentation: pop-ups, rich text, ...
- Python GUI toolkits are multiplatform
- Make it accessible for non-Greybeard

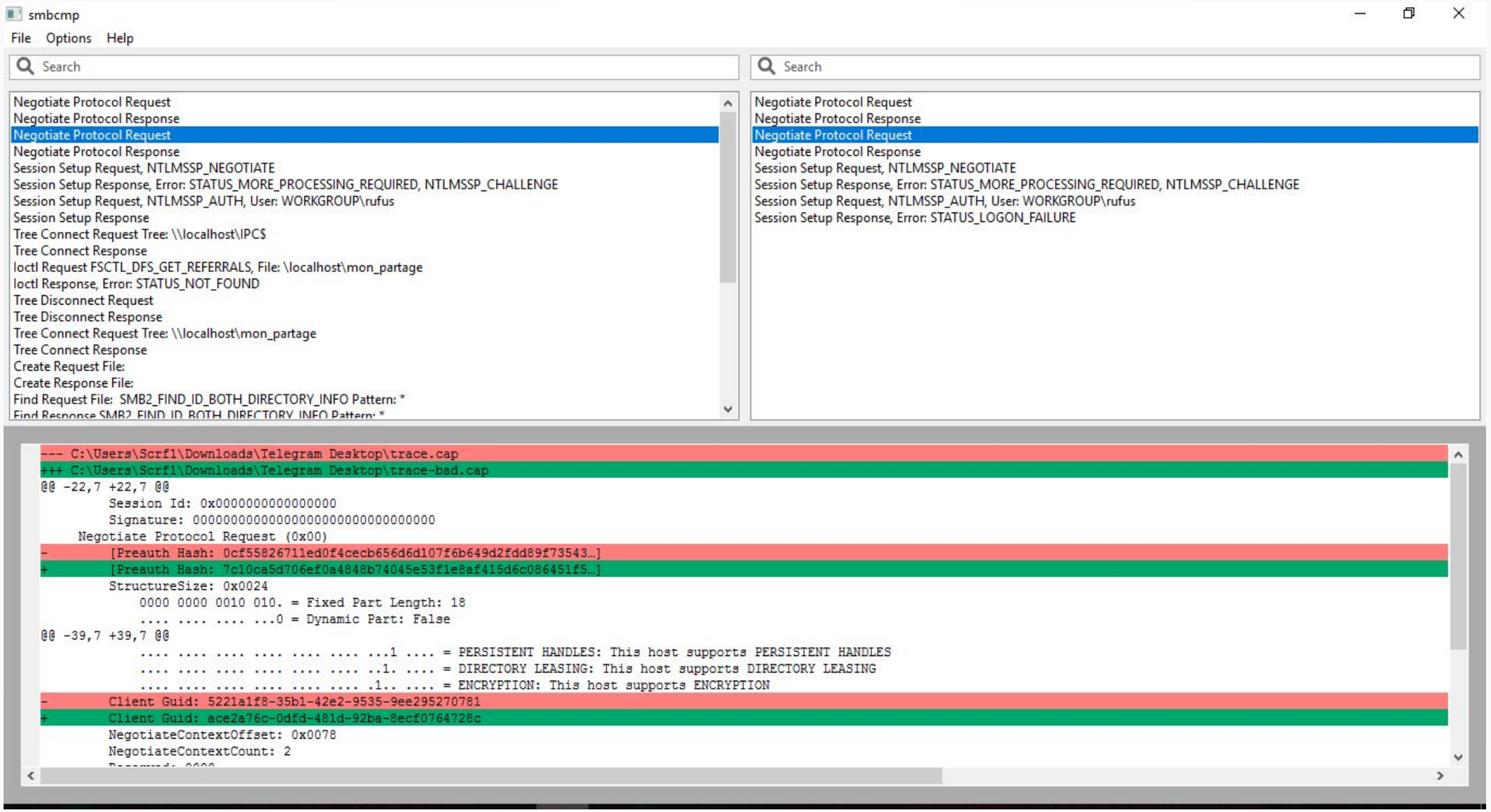
Why WxWidgets?

Framework	License	Documentation	Wysiwyg	Target	Native
WxPython (Phoenix)	WxWindows Library License (~LGPL)	Good	Yes	Desktop	By default
Tkinter	BSD	Good	No	Desktop	Painful
Pyside 2 (QT for Python)	LGPLv3/ GPLv2/ Commercial	Poor	Yes	Desktop	Painful
PyQT	GPL/ Commercial	Good	Yes	Desktop	Painful
Kivy	BSD	Good	No	Mobile	No
PyGTK	LGPL	Medium	Yes	Desktop	Only on Gnome
PySimpleGUI	GPL v3	Good	No	Desktop	Yes

Plus it looks good on Linux (Gnome)...



And Windows


smbcmp
File Options Help
Search
Search
Negotiate Protocol Request
Negotiate Protocol Response
Negotiate Protocol Request
Negotiate Protocol Response
Session Setup Request, NTLMSSP_NEGOTIATE
Session Setup Response, Error: STATUS_MORE_PROCESSING_REQUIRED, NTLMSSP_CHALLENGE
Session Setup Request, NTLMSSP_AUTH, User: WORKGROUP\rufus
Session Setup Response
Tree Connect Request Tree: \\localhost\IPC$
Tree Connect Response
Ioctl Request FSCTL_DFS_GET_REFERRALS, File: \localhost\mon_partage
Ioctl Response, Error: STATUS_NOT_FOUND
Tree Disconnect Request
Tree Disconnect Response
Tree Connect Request Tree: \\localhost\mon_partage
Tree Connect Response
Create Request File:
Create Response File:
Find Request File: SMB2_FIND_ID_BOTH_DIRECTORY_INFO Pattern: *
Negotiate Protocol Request
Negotiate Protocol Response
Negotiate Protocol Request
Negotiate Protocol Response
Session Setup Request, NTLMSSP_NEGOTIATE
Session Setup Response, Error: STATUS_MORE_PROCESSING_REQUIRED, NTLMSSP_CHALLENGE
Session Setup Request, NTLMSSP_AUTH, User: WORKGROUP\rufus
Session Setup Response, Error: STATUS_LOGON_FAILURE
--- C:\Users\Scrf1\Downloads\Telegram Desktop\trace.cap
+++ C:\Users\Scrf1\Downloads\Telegram Desktop\trace-bad.cap
@@ -22,7 +22,7 @@
Session Id: 0x0000000000000000
Signature: 00000000000000000000000000000000
Negotiate Protocol Request (0x00)
- [Preauth Hash: 0cf55826711ed0f4cecb656d6d107f6b649d2fdd89f73543…]
+ [Preauth Hash: 7c10ca5d706ef0a4848b74045e53f1e8af415d6c086451f5…]
StructureSize: 0x0024
0000 0000 0010 010. = Fixed Part Length: 18
....0 = Dynamic Part: False
@@ -39,7 +39,7 @@
....1 .... = PERSISTENT HANDLES: This host supports PERSISTENT HANDLES
....1. .... = DIRECTORY LEASING: This host supports DIRECTORY LEASING
.... .1.. .... = ENCRYPTION: This host supports ENCRYPTION
- Client Guid: 5221a1f8-35b1-42e2-9535-9ee295270781
+ Client Guid: ace2a76c-0dfd-481d-92ba-8ecf0764728c
NegotiateContextOffset: 0x0078
NegotiateContextCount: 2

Supported platforms: Linux

- Works out of the box
- Wireshark CLI (Tshark) needs to be installed
- Optional dependencies:
  - LXML: faster than (c)ElementTree for our use case: lxml.de/performance.html
  - Wxpython (for the GUI)

Packaging for rpm based distributions

- Difficult because each specfile has different guidelines
  - Fedora: docs.fedoraproject.org/en-US/packaging-guidelines/
  - Opensuse: en.opensuse.org/openSUSE:Specfile_guidelines
- Need to package all the dependencies not already packaged
- Very tedious

Supported platforms: Windows

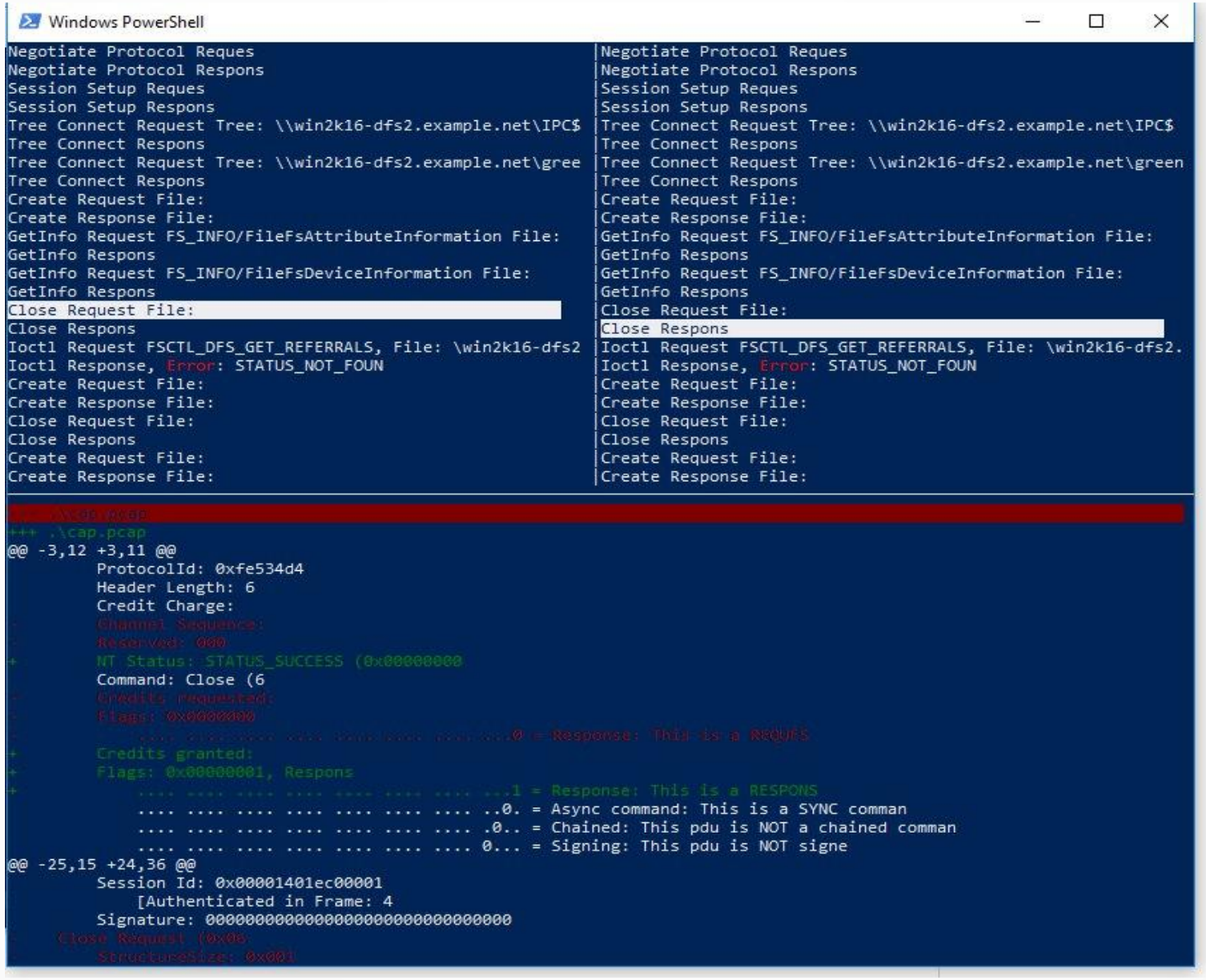
- The GUI works out of the box
- The CLI needs tweaking: Cygwin, Powershell, WSL

Port the CLI to Windows

- Bundle a wireshark build stripping useless things
- Bundle a Python build (embeddable)
- A C program launches the Python interpreter with correct arguments to start smbcmp

Final result:
github.com/smbcmp/smbcmp/releases/download/v0.1/smbcmp-x64-0.1.zip

Final result on Powershell



Supported platforms: macOS

- It works, but it hasn’t been tested (TM)

In retrospective

- GSOC was a really good experience
- email-based open source development (bazaar) was weird and seemed unnatural
- My mentor was great and always available
- The imposter syndrome is real

  Final work submission: rmpr.github.io/gsoc_2019/

Time for a little demo...

Follow-up

Qtwirediff

github.com/aaptel/qtwirediff

- **Experimental: Generalization of smbcmp to every protocol**