

GCIA Practical Study and Planning Guide 4.2

This document was prepared by Jeff Holland (GCIA, GCUX, GCIH, GSEC, CISSP); Jamie French (GCIA, GCUX, GCWN, CISSP); and Koon Yaw Tan (GCIA, GCUX, GSEC) to assist other GCIA candidates in preparing for and passing their certification. Many thanks to Jeff, Jamie, and Koon Yaw for their contribution to the community! If you have suggestions for changes or improvements to this document, please send them to lara@sans.org with the subject line "GCIA Study Guide".

The GCIA practical is a significant task, both in terms of difficulty and time. The following is a guideline meant to help you plan and stay on track so that you may successfully complete the practical on time, and hopefully obtain a passing score of 70% or better.

While work emergencies arise when we least expect them, as well as family and other responsibilities being a factor, you should plan to start your practical as soon as possible. The biggest threat to your completion is procrastination.

The next threat is not following the instructions! **You MUST read, understand, and follow both the Assignment and the Administrivia.** If they call for a 10-page minimum, be sure your paper is at least 10 pages long. Pay particular attention to correctly citing references and giving sources the credit they are due. If you have any questions concerning references, the Administrivia document contains a link to sample references so you know how to use/format them. If you have questions about a practical requirement (for example, the link graph), refer to your course material and/or the Intrusion Detection Forum at <http://forum.sans.org>. If you still have questions please post to the Intrusion Detection Forum or write to info@sans.org for more direction. SANS and/or the GCIA Advisory Board will do their best to answer your question in a timely manner. **If your practical does not meet the requirements set out in the instructions, it will not pass.**

Note to self-study/online students: You must carefully plan your six-month timeframe to complete the modules/quizzes, practical, and study for the exam. Since you do not have the luxury of having finished the course work at a conference, you might consider giving yourself 12 weeks to finish the modules and 12 weeks to finish the practical and exams. Whatever you choose to do, plan to give yourself at least 10 weeks to do the practical and finish the exams.

Note to conference students: While you have the benefit of having the course work behind you, you must still plan carefully to finish the practical on time and study for the exams. It generally takes about a week after the conference before receiving access to your account. You may use this time to work on your real/theoretical test network. If you do not have an IDS setup, this is an opportune time to wrap yourself around the task and chose one (either commercial or free) and install it. You have 5 months to complete the practical, and an additional four weeks to take the exams. You may finish the practical early if you wish to give yourself more time to study for the exams.

Note to challenge students: Since you do not have the complete courseware, you will have to plan your time for your self study. Practical assignments are also updated periodically, usually about 3 times a year, and they do have expiration dates. Check the assignment updates occasionally to ensure you do not miss a deadline for a retired assignment. Do not combine two versions of the assignment. For this reason, you must carefully plan your time. For more details on the challenge option, refer to http://www.giac.org/cert_programs.php.

Whichever the case, it is advised that you take a few minutes to review the entire practical requirements before starting. This will allow you to determine how much work each section might take, and also what skills you will need to study up on to complete the practical.

Practical Assignment

You have been asked to provide a security audit for a corporation by analyzing honey-pot alerts logs from their intrusion detection systems and security point devices to produce an analysis report. You have also been asked to analyze one raw detect/attack from a separate network using raw logs. The practical consists of four parts. In the first part, you are required to write an executive summary. The second and third parts contain the meat of your practical, the honey-pot and raw log analysis. The report will then be concluded with a description on how you arrived at your results. The executive summary will obviously be aimed towards management, so you should take extra care in how you present your information. Parts two, three and four will most likely be aimed toward other intrusion analysts.

The practical is worth 100 points in total, and is distributed among the different sections (or “Parts”), as follows:

Part I - Executive Summary (10 Points)

Executive summaries help bridge the gap in technical expertise between you the technical person, and management, who have to make informed business decisions. This is a critical job function for people employed in a technical capacity. You have to clearly and concisely communicate what the main areas of interest are according to what you expect management must address and management will be relying on you to do this well. The management should be left a clear view of the overall security health of their organization (based on your observations). The executive summary should not exceed two pages of text, but you must still make sure that you convey the most important points. Keep the objective in mind while writing. Remember, this is a summary.

Part II & III- Detailed Analysis (45 Points and 25 Points, respectively)

This part is worth 70% of the practical and you should spend most of your time making sure your analysis is accurate and meaningful.

There are many different ways to approach data mining in support of analysis. Some solutions are OS/platform specific, however the majority of solutions may be used cross-platform. For instance sed, awk, grep are generally accepted as UNIX tools but are available on Win32 platforms as well. Another common example are tools that use PERL or even PHP. These too have been ported to Win32 platforms and allow scripted applications like SnortSnarf to run under Windows. On a similar vein, scripts written in VisualBasic and written specifically for Win32 may not run under UNIX.

It is up to you to decide upon which platform you use for data mining. It is recommended that you choose the platform you feel most comfortable using on a daily basis. Whether you choose UNIX or Win32 or some other platform, a certain degree of configuration will be required in order to get the tools operational anyway. If you are using someone else's assets (like your employer's) make sure you obtain permission to do so first.

Here are some common data-mining solutions and a generic list of the tools used. There is no "right" or "wrong" way to approach the problem, there are pros and cons associated with each. Choose a solution that balances the useful information you can gain from the analysis with your comfort level with particular tools or techniques. Deciding upon your analysis process and setting up your systems to perform analysis should take no longer than two weeks.

Analyze the logs independently as separate files, based on date:

Pros: 1. Prep-work to get working on analysis is minimal

Cons: 1. Analysis conducted tends to be near-sighted
2. Many of the same alerts or events of interest need to be analyzed again (each day).
3. Papers done this way tend to have issues with how they flow, are much harder to read, as well as present challenges in formatting
4. Overall effort within the analysis sections is multiplied by the number of days analyzed and usually does not do a good job of relating data from one day to the next

Analyze the logs inclusive of dates selected using command line tools:

Pros: 1. It isn't difficult to concatenate all the files together based on log type
2. Provides for flexibility in how the logs will be analyzed
3. Does not require in-depth knowledge of scripting or writing code
4. Papers done this way tend to be general in scope but do an adequate job of identifying detects that are pre-determined as being ones the student wishes to cover

Cons: 1. Some basic knowledge of the commands being used is required (or search posted GCIA papers on giac.org for how other student have done this)

2. Not knowing how to write batch files or simple scripts may make analysis awkward
3. Relational analysis is difficult without strong analytical skills

Analyze the logs through scripts, either canned or custom:

- Pros:
1. Someone else has probably already made the script for you
 2. Scripts can be modified to suit the task at hand without having to rewrite them from scratch
 3. Analysis can be relational with both simplistic and complex relationships identified
 4. The output can be customized, sorted etc.
 5. Scripts are portable (PERL will work on both Win32 and Unix)
- Cons:
1. Knowledge of the scripting language is required if customizations are needed
 2. Installing the interpreter can be intimidating
 3. Analysis tends to be flat and lacks relational depth, becoming more statistical in nature

Analyze the logs through the use of a database:

- Pros:
1. Relational analysis becomes much easier. Important detects are much more easily identified.
 2. Data management becomes much more transparent to the student.
 3. Analysis process is much more efficient
 4. Output from the database queries can be very meaningful and queries highly customized to retrieve complex relationships
- Cons:
1. Setting up a database and importing the information can be daunting to many people
 2. A general understanding of the Structured Query Language (SQL) is needed
 3. The student should know basically what to look for prior to beginning analysis (usually not good for beginners in the IDS field)

Windows Tool Resources:

| | |
|------------------|---|
| awk | http://www.muc.de/~walkerj/GAWKDLL/gawkdll.htm |
| ConText | http://www.fixedsys.com/context/ |
| Cygwin | http://www.cygwin.com/ |
| Grep for Windows | http://www.interlog.com/~tcharron/grep.html |
| MySQL | http://www.mysql.com |
| PERL | http://www.cpan.org/ports/ (recommended) |
| PERL | http://aspn.activestate.com/ASPN/Downloads/ActivePerl/ |
| sed | http://www.student.northpark.edu/pemente/sed/#ssed |

| | |
|------------|---|
| Snort2Html | http://www.geocities.com/swan_daniel/snort2html.txt |
| SnortSnarf | http://www.snort.org/dl/contrib/data_analysis/snortsnarf/ |
| Snort-Sort | http://www.dpo.uab.edu/~andrewb/snort/snort_sort.html |
| Textpipe | http://www.crystalsoftware.com.au/textpipe.html |
| TEXTTools | http://www.FireflySoftware.Com |
| Wingrep | http://www.wingrep.com/download.html |
| WinGrep | http://www.mwso.com/eng/wgrep1.htm |

Unix Tool Resources:

(most command line tools used for manipulation of data come with *NIX distributions by default)

| | |
|------------|---|
| MySql | http://www.mysql.com |
| PostgreSQL | http://www.postgresql.org/ |
| Snort2Html | http://www.geocities.com/swan_daniel/snort2html.txt |
| SnortSnarf | http://www.silicondefense.com/software/snortsnarf/index.htm |
| Snort-Sort | http://www.dpo.uab.edu/~andrewb/snort/snort_sort.html |

Some good reference practicals:

Berkley Database

<http://www.zeltser.com/sans/practical/#a3-1>

AWK, sort scripts

http://www.giac.org/practical/Chris_Calabrese_GCIA.zip

MySql Database, sed and awk

http://www.giac.org/practical/Brandon_Newport_GCIA.zip

http://www.giac.org/practical/Jason_Lam_GCIA.doc

PERL sorting script

http://www.giac.org/practical/Bill_Phillips_GCIA.zip

PERL scripts

http://www.giac.org/practical/chris_kuethe_gcia.html#3.0

PERL sorting script, grep

http://www.giac.org/practical/dana_mclaughlin_gcia.doc

Unix shell scripts, awk, grep

http://www.giac.org/practical/Chris_Baker_GCIA.zip

Unix shell scripts, awk

http://www.giac.org/practical/PJ_Goodwin_GCIA.doc

Windows PERL, grep

http://www.giac.org/practical/Loras_Even_GCIA.doc

Windows batch scripts, TEXTTools

http://www.giac.org/practical/John_Topp_GCIA.doc

You should have a good idea of the approach you wish to take, and have already setup your work environment so that you are ready to start. The first thing that should be done with the data is to “organize” it. There are many, many ways to do this. Some common methods are to sort by frequency of occurrence by event, source IP address, destination IP address, source port and destination port. Others sort their logs by priority, meaning that an alert that does not have a high quantity of log entries might be more severe and deserves a higher priority. Be creative here but make sure you cover the logs adequately. The best practical submissions contain a combination of ordering, for example they will be ordered by destination IP and by severity. **It is important to define your scope before you start. Figure out your overall analysis process and stick to it.** This will greatly aid your overall analysis’ flow and keep it smooth and focused. The less organized you are, the more chaotic and unfocused analysis tends to be, which negatively impacts the results.

Now that you have a plan of attack, you will need to do some research. Correlate the events of interest from the log files with external sources. This should include relevant postings on the Internet, paper publications, and other student practicals, as well as with other related events of interest from within the log files themselves. Make sure you understand what caused the event stimuli. This greatly focuses on the Snort rules themselves, as it is the underlying log generator. If you do not understand what triggered the log event you may be building your analysis on a shaky foundation.

You can probably guess it by now. You are set up to do some relational analysis and reinforce much of the material taught on the SANS GCIA track. Remember to look at related events of interest when performing analysis. This is where the most interesting information and probably the most critical analysis will be completed.

We will now cover the assignment requirements step-by-step to help clarify any questions.

1. *The files you choose to analyze. If any file types are skipped, points will be deducted and your analysis will be considered incomplete.*

It is not hard to get a good score for this requirement. List all the files you analyzed.

2. *Meaningful analysis identifying relationships between the different computers that generated logs. **This will provide a basis for the overall grading of this***

assignment. Take note of the text in bold! Relational analysis is crucial in arriving at informed conclusions and is almost always at the core of in-depth analysis. If you do not cover all your bases and try to piece together the big picture you will almost assuredly be missing out on something you ought not to. The score for this requirement is assessed after the assignment is finished being graded and reflects the overall impression the grader had of your effectiveness in finding and analyzing the events of interest within the complete data set. For example, when you analyze events in great depth, do not choose 2 related to the RPC DCOM based worms and one that turns out to be 99.9% a false positive where the 0.1% of true positives were benign reconnaissance probes. Use some logic in your prioritization and try to cover the main and most critical events of interest affecting the assets on the network. These are usually the assets that have the highest severity or assets that would cause the largest impact if they fell under control of the WRONG people.

This is an overall mark assessed on the quality of the analysis. If the student identified an event of interest but did not discover related events based upon their original analysis their score will be adversely affected. This also means that reiterating what a Snort rule is supposed to trigger on is also not adequate. It is expected that the student progresses beyond this and tries to figure out if this event of interest was a false positive, indicative of a network problem, was an actual compromise, or deserves any defensive adjustment considerations.

The link graph must be presented in a graphical format. It does not have to be fancy and full of color but does have to highlight a relationship that would be obscure to someone without the aid of a graphical representation. Be creative here. Obviously, the stronger the relationship identified, the more functional the link graph.

3. *An overview of what detects you have identified, and the number of occurrences of each detect. From these, select the three most critical detects and perform an in-depth analysis. Please note that the most noisy detects are not necessarily is the most critical ones.*

List out the suspicious detect and select three detects that you deemed critical (two from the honey-pot logs and one from the raw logs). Remember to provide reasons why you select them. What is sought here is actually the meat of the analysis. Did you cover the detects of interest you decided to present upon adequately? When you have a detect, it should be expanded upon through research and cross correlation with other log entries. Answer all the questions related to the detect here. Was it a false positive? Was this the only event the attacker triggered? Is there any corresponding activity from the organization that indicates compromise around this time frame? If you answer the “who, what, when, why, where, and how” with appropriate support you

will score very well here and it will ripple through the rest of the assignment, bolstering your mark overall.

Review the following paper concerning this part:

http://www.giac.org/ID_assignment_guidelines.php

Pay special attention to the intent of the spoof probability and correlation sections. The spoof probability section should address how strongly you believe the source IP was spoofed, and why or why not. The correlation section should address other reported instances of this attack, or state that the attack has never been seen before. A correlation to the attacker's IP address to determine how prevalent that IP address has been in recent attacks would also be a nice addition to this section. In other words, has this IP been reported to other sources as an attacker? Check the www.dshield.org site or use a search engine such as Google. If there were no other correlations to this attacking IP, state so in your write up and include correlations of similar attacks. It is also a good idea to add links to [CVE](#), [CERT](#) or other advisories in this section. Also, be sure to give reasons for your scores in the severity equation. Reviewing your GCIA coursework as to what each of the categories means and tries to identify is a good idea.

Here are some links that can help you. Keep in mind that this list is by no means exhaustive.

<http://www.SANS.ORG/newlook/resources/IDFAQ/oddports.htm>

<http://www.iana.org/assignments/port-numbers>

http://www.switch.ch/docs/ttl_default.html#overview

<http://www.whitehats.ca/main/tools/portquery2/portquery2.html>

<http://www.cert.org/>

<http://www.cve.mitre.org/>

4. *Different "network statistics" consisting of:*

- *A "top talkers" list presenting who the top five talkers are in terms of the log files analyzed. You may use different criteria to make your list as meaningful as possible (Be sure to define the criteria you used to select the "top talkers").*
- *A list of the five top targeted services or ports. Explain your criteria and - if possible - what exactly you think is being targeted (it does not have to be HTTP just because it is port 3128).*
- *A profile of the three most suspicious external source addresses. You must also include the reasons why you selected those hosts for further investigation. Your profile should at least include registration information, but you are encouraged to supply any other information you have concerning them (for example, you may be able to take a guess on operating systems). Be sure to*

read the policy regarding active scanning at the end of this document before you do anything that will get you in trouble.

This is to understand who are the top attackers, victims and targeted services. You should have adequate discussion for the listed top talkers and targeted services. Simply tabulating them without any explanation will not lead you a passing score.

We also want to test your ability to determine who the attacker is. This is very helpful in resolution and analysis in general. It also puts a name to the objectionable traffic and in many instances really helps in understand what is going on. Names give away a lot of information. We are trying to express this through practice so that students can take this back to their place of employment and have this basic skill set so they know how to contact the appropriate individuals when an incident arises. List the registration information for three selected IP addresses and state why they were chosen.

- 5. Correlations from previous students practicals (GCIA #0600 and above) and/or from other sources. Be sure to appropriately cite and credit the sources used for correlation. Graders are looking not only for correlation with other students papers which will very likely be directly related as it deals with the same type of events on the same network, internal self referencing correlation with other events and logs, as well as external correlation from other sources. If for example you do not reference other students previous practical assignments where relevant but you do a fantastic job of assessing the events being discussed you will still receive a passing mark. It is strongly recommended however that you review previous students posted works for correlation. You should also correlate your analysis with current trends from the Internet Storm Center (<http://isc.sans.org>) and similar sites whenever possible.*

Not surprisingly, this is listed as a requirement. It is done to emphasize to the student the value of looking at some other assignments. A lot of coverage has been completed on very similar detects within other students' practical submissions. It is highly recommended that these be used for correlation purposes because they are focused on the same network environment and would be the best source to consult for correlation in almost all cases. It is not wrong to expand upon others work, as long as sufficient credit is given when their work is used. Other external correlation is also sought here where correlation from other students' practicals is not enough. Points will be deducted if there is insufficient correlation to support your analysis, whether it be from external source or other students submissions.

- 6. Any insights into internal machines such as compromises or possible dangerous or anomalous activity.*

Many students cover this requirement within their analysis. We are looking for the student's conclusions from analysis here. Take a stand. When you see something that looks wrong, dangerous, anomalous, or like a sure fire compromise indicate this and flag it for some defensive recommendations. Including a summary of these machines at the end of your assignment is most helpful as well. Failure to identify problem machines (unless analysis indicates there were none) will adversely affect this score.

7. *Defensive recommendations based upon your analysis.*

This is the product of analysis that we have been striving towards. Now that we have completed analysis we should sit down and figure out how to correct any problems and recommend them to the appropriate authority so that we can become proactive vice reactive. It makes sense that when you notice something broken you will not only try to fix it, but also fix what caused it so it doesn't happen again. Defensive recommendations need to be focused on the analysis conducted within the assignment. Listing generic defensive recommendations that do not specifically address of the corporation will be awarded a penalized score. Make the recommendations relevant to the corporation and based on your analysis for best marks.

Part IV - Analysis Process (20 Points)

Here you get to brag about how you tackled the logs. What you did to organize the data. How you decided to categorize, organize, and prioritize logs. What custom scripts you may have used. Which tools you found to be effective. What new commercial product helped you. What new methods of using old tools or generally non-related tools were invented to help. Did you develop new ones to do the job for you? This does not have to be lengthy and can be a simple description of how you used grep and awk. It could also be very technical and in-depth, including SQL statements used on your database, describing the process of how a program handled the data, problems found while trying to get SnortSnarf to run etc. This gives the grader an understanding of your methodology and also provides tips and tools for other students to use or expand upon your process.

You should include a brief list of the software (including OS) and hardware you deployed to tackle the logs. If you ran into problems with your tools, document these and your solutions or workarounds. As appendices do not count towards the maximum length requirement, you are welcome to include any scripts or programs you have written (if your scripts are very long, you may wish to link to them hosted elsewhere so that others may benefit from them without filling up the report).

Finalizing Your Practical

You have put many days of hard work into your practical. Before submitting your practical, spend some time to review for typos and missing details. It will be quite difficult to catch the small errors as you have by now been staring at the practical for

many straight. Also use this time to review the practical assignment again and make sure you have not missed anything. It is imperative that you make sure all external sources of information and reference are properly cited. SANS has a low tolerance for plagiarism. Not only could your paper be failed for failure to properly cite sources but you could also be banned from taking GIAC certification for a period of time; in severe and/or deliberate cases, you may have previous certifications revoked. This is a serious issue! Make sure you properly cite your resources and give credit to those that deserve it.

Now, have a friend or colleague review your paper. This person should be looking for mistakes and also be able to clearly understand what you are presenting. If there is something they don't understand then perhaps it needs to be modified. Chances are that if they are confused others will be too. Check your practical with a spell checker and also have your reviewer check for spelling and grammar errors. **Once submitted you will not be allowed to make modifications prior to posting on the Internet where everyone will be able to see your mistakes. Neither graders, nor the webmaster, will correct these errors prior to posting. Be sure you have thoroughly reviewed your final copy before sending it! Make sure *all* sensitive material is sanitized, as well.**

Additional Notes:

If you wish to achieve honors status on your practical, your practical needs to go “above and beyond” the minimum requirements. Your paper should be well laid out and have an excellent presentation effect. If you do original work, this will help push your paper into honors territory. Displaying a strong understanding of the course material in your work will also aid in making your practical a candidate for honors. Be sure to review other honors papers and their quality. You should strive to achieve the same or better quality than those papers display.

If you have not noticed already, following the practical assignment instructions has been stressed repeatedly. **If a practical does not follow the instructions, it will not pass.** Do yourself and the graders a favor...follow the instructions, write a great paper, and help contribute to the security community in a positive way with your work. Good luck!!

- The GCIA Graders and Advisory Board