Distributed WPA Cracking

Rodney Beede

University of Colorado  
Boulder, CO 80309-0430



Ryan Kroiss

University of Colorado  
Boulder, CO 80309-0430



Arpit Sud

University of Colorado  
Boulder, CO 80309-0430



**ABSTRACT**

In this paper, we describe a distributed system we developed for doing lookups in a rainbow table for passwords to WPA-PSK 1/2 wireless encrypted networks. Our motivation for developing such a system came from the abundance of distributed systems for generating rainbow tables but a lack of any to do the lookup.

The rainbow tables for even just a million passwords can be gigabytes in size. The problem we address is how a distributed system could be used to provide fast lookup of matching passwords in a table for some given captured wireless network data. We chose to focus on WPA versus other wireless encryption techniques (e.g. WEP) because it offers the best encryption that cannot currently be feasibly defeated via brute force methods.

Our approach involved writing new code to handle user submitted jobs of wireless data packets and modifying existing code from a project known as coWPAtty [5] which handled the actual table lookup. We used a cluster of 9 nodes running on virtual machines to handle the work load and measured the performance of our system to the serial original coWPAtty implementation.

Our results show an order of magnitude of 8 times in speed increase for our distributed system versus the serial coWPAtty. Testing data showed that the original serial coWPAtty code could still return a single result in an order of several seconds while our system returning results in under 1 second. While not a seemingly big difference the scalability of our system would provide much more usability to offer this as a service.

We concluded that our results show that the importance of using a strong password for wireless networks is still very important. WPA is an improvement over previous encryption ciphers used in wireless networks, but it is still susceptible to weak passwords chosen by the end user. The most important outcome from our research shows that using large rainbow tables of pre-computed passwords can provide an easy, fast, and scalable method for finding weak passwords in encrypted wireless networks.

**Categories and Subject Descriptors**

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2011

Template - Copyright 2010 ACM 1-58113-000-0/00/0010

C.2.1 [**COMPUTER-COMMUNICATION NETWORKS**]: Network Architecture and Design– *Wireless communication;* K.6.5 [**MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS**]: Security and Protection– *Unauthorized access (e.g., hacking, phreaking);*

**General Terms**

Security

**Keywords**

WPA, rainbow table, dictionary attack

# INTRODUCTION

The use of wireless networks based on the 802.11 standards (also known as WiFi) has become much more common place in your typical household. Early advents of these networks provided security to limit access and protect sensitive data with a protocol standard known as WEP. Research and analysis into the cipher algorithms used in this protocol lead to discovered weaknesses that could easily be exploited by a single computer with modest hardware using brute force. The deficiencies of this protocol lead to the development of a new one known as WPA [3]. This newer protocol utilizes much more advanced and stronger encryption ciphers which have made the possibility of using brute force attempts not feasible [4]. With the advent of these new security measures ways of attacking them have evolved as well.

One technique for deciphering encrypted data or accessing protected resources is to simply try a large number of different passwords in the hope to get one correct. The nature of wireless networks enables one to capture encrypted packets as they pass through the air and store them to disk in their encrypted form. Part of the wireless encryption protocol involves handshaking at the initial connection in order to authenticate the client to a wireless access point. This handshake includes some initialization vectors (IV’s) that are used to allow the client to encrypt the password they will send for authentication. There are multiple types of WPA encryption some which use client certificates and server certificates for encryption of the data while others use simple passwords (WPA-PSK) that the user provides upon establishing the connection. The password has a minimum of eight characters and a maximum of 63 and is case sensitive. The encryption key is derived from this password and the case sensitive name of the wireless network (SSID). Attempting to brute force this key space would also prove to be computationally unfeasible. A common alternative is to try a large dictionary of common passwords. This dictionary can also include slight permutations on words such as appending numbers to the end in order to catch common passwords used by users.

The dictionaries used could contain millions of words which would require some significant amount of time to process for each wireless network someone wished to attack. A complimentary technique is to generate a large rainbow table of keys based on the dictionary and several common wireless network names. This would then enable an attacker to simply capture some wireless network handshake data (IV’s) and do a lookup in the pre-computed rainbow table for a match. Such a tool already exists and is known as coWPAtty.

coWPAtty is a serial program that can generate a rainbow table with hash values for matching keys based on wireless networks names (SSIDs) and WPA-PSK passwords. The generation is done in a serial manner and written to a file that can later be used in conjunction with captured wireless data. All operations in the original coWPAtty code are done in a serial manner on one machine.

For our project we decided to create a distributed system that could perform this rainbow table lookup among a cluster of nodes in order to increase performance. We chose to not generate the rainbow tables as others have already done so [2]. Instead we chose to use an existing rainbow table that contained pre-computed keys for 1,000 wireless network names (SSID’s) and divide it across multiple machines for doing lookup queries.

We developed a system with a single master node and multiple worker nodes that handled the job submission and work. The master node was written from scratch as a Java web application and is responsible for queuing jobs, sending them to the workers, and reading back the results. The worker nodes were created by modifying the original coWPAtty code to function in a distributed manner and are described later.

Our testing methodology consists of capturing wireless data from our own personal networks and submitting them to the distributed system. We compared the times for our system to find the correct solution to that of the original coWPAtty serial code running on a single node. We document our findings in the results and conclusions portion of this paper.

All the project code and testing data can be found on the project’s website at http://code.google.com/p/distributed-wpa-cracking/

# ARCHITECTURE

## Overview



Figure 1. Architecture Overview

The system is comprised of a master node which is responsible for coordination of work among a cluster of worker nodes. The worker nodes each hold a portion of the rainbow table in memory and listen for requests over TCP from the master node.

A network file system is used as a shared location for the job input data and output results. In addition it serves as a central point for the binary code that will be run on both the master and the worker nodes.

The wireless network is not connected to the master or worker nodes nor must it even be anywhere in proximity to them. The end user captures wireless data from the wireless network using their computer and saves the result to disk. At a later time, after sufficient data has been captured, the user can then upload the captured data to the master node for processing.

Communication from the wireless network to the user’s computer is done via standard 802.11 wireless protocols. The data capture can be done with already existing tools such as Aircrack-ng [1]. Communication from the user to the master is accomplished using the HTTPS protocol over a secure TCP socket. Communication internally occurs between the master node to a worker node. Worker nodes do not communicate with each other. The master node uses TCP connections with the SSH protocol for remotely starting the worker node binaries and a custom propriety protocol for sending running worker nodes jobs and checking their status.

## Network File System

In order to be able to share input data (the wireless capture file) with the worker nodes a networked file system was setup with common access between all worker nodes and the master node. This choice was made because it was a simpler configuration versus having the master transmit a copy over TCP to each worker node and have them store a temporary copy on their local disks. The master node was used to host the network file system, but a separate device could have been used as well.

When the user uploads a job request to the master it stores the uploaded capture file to a newly created folder on the network file system. The folder is given a name that matches the newly generated job id. This folder is visible to the worker nodes and will be used by them to later read the wireless capture file and also for storing the output results.

Common binaries for all the workers such as the actual worker binary executable are stored here as well. This allows for updated code to be published in one place and be applied to all workers at once. A simple restart of the worker process on each worker node will then load the new updates to any code.

The large rainbow table data however is not stored on the network file system but instead a read-only copy exists on each node on local disks. Because gigabytes of data must be read from the rainbow table by each node at startup into memory having copies on local disk allow for shorter startup times since disk contention and network congestion are avoided. While this would add some burden to adding updates to the rainbow table this isn’t done frequently and just requires more time to copy any new data to each node’s local disk. In addition our project doesn’t generate new rainbow tables so no updates were really necessary.

## Master Node

The master node is a Java web application written from scratch that is responsible for accepting and queuing user requests along with tracking the status of the worker nodes.

The master node web application provides a user interface as depicted in figure 2 that provides the user with the following:

* A HTML form for uploading a new job
* View of the jobs in the queue and completed jobs
* Status list for the worker nodes
* Controls for starting / killing the worker nodes
* View of the web application log



Figure 2. Master Node Web Application Interface

### Configuration

All configuration is managed via an XML file on the master node that is loaded at the startup of the master node web application. It specifies the following:

1. The base directory pathname for where job data should be stored
   1. A network file system share that is accessible to both the master and worker nodes
2. The hostnames and port addresses for the worker nodes
3. The directory to the rainbow table for local use
4. Command templates for starting the worker nodes
5. Command for killing worker nodes
6. SSH credentials for accessing worker nodes

The template approach for the start command for worker nodes allows for the master to fill in necessary details such as directory paths and rainbow table offsets when starting a worker node. Details will be discussed in the section on worker nodes.

### Job Submission

The web site itself is protected with a username and password along with HTTPS encryption. When a user submits a job via the interface the username, as reported by the web server container (Apache Tomcat), will be used for the name of the job owner and as part of the job id.

The user simply uploads a wireless network data capture and specifies the SSID of the desired network (a data capture could have more than one network in it so this must be specified). The upload and job submission could also be accomplished through automated means of a script by using the standard HTTP protocol as well.

The master node web application receives the request, generates a unique job id for it, creates the necessary directories on the network file system, and adds the job to the queue.

### Job Queue

The queue is served in first-in first-out order and a history of all jobs is retained until the master node web application is restarted.

A job will not be started until all workers report that they are in a ready state and not currently busy with another job. Jobs are run one at a time.

Information such as the start and end time along with the solution found, if any, is listed in the interface. All users can view any job in the queue.

When the master it told that one worker has found the solution it will send a kill job signal to all other workers informing them to stop in case they haven’t yet.

### System Status

The master periodically queries each worker to get their current state. The list of workers and their state is listed to the user. Details on the various state types are listed in the Worker Node section later.

### Start / Kill Worker Nodes

The master web application is responsible for initially starting each worker node and also provides a mechanism to terminate them all. A worker node is only started once before any jobs in the queue are processed. Each job that starts does not retrigger the start of a worker node. The master uses an SSH connection to remotely connect to the worker node specified by the configuration and issues a command to start the worker. The executable is accessible via a network file system share, and an appropriate pathname is used in the command to this binary executable file.

Before starting all the workers the master also verifies that the rainbow table SSID’s are all consistent (the same size). It also calculates the byte offsets that each different worker node will use when they are to load their portion of the rainbow table.

## Worker Node

The worker node is a modified C program based on code by Joshua Wright’s coWPAtty. It has been modified to act as a type of service that listens on a TCP socket for instructions or status queries from the master node. The communication specifics will be discussed in the “Master to Worker Node Communication” section later.

### Worker States

A worker node may be in any of the following states which are reported to the master node:

* NOT LOADED – the worker is not accepting TCP connections. It may not be running at all or it may still be busy loading the rainbow table into memory.
* LOADED – the worker is ready to accept jobs, has loaded the rainbow table, and has not yet run any jobs.
* RUNNING – the worker is currently busy running a job. It will refuse to run any additional jobs and return an error if asked to do so.
* FINISHED – the worker has finished a job and is ready for the next.
* ERROR – the worker has encountered an error and can no longer run any jobs. It must be restarted.

### Rainbow Table

The master node web application will provide via command line arguments to each worker node the relative offsets each should use for the beginning and end of the rainbow table data. In addition a command line option of where to find the rainbow table files is provided. A portion of each network SSID from the rainbow table file is loaded in an even distribution across all workers. This allows them to divide up the work of searching the rainbow table to reduce the time a lookup takes. In addition the entire rainbow table portion is loaded into memory so that performance is not bottlenecked by the disk.

### Computation

When a worker node receives a new job from the master, it first checks that no other jobs are running. An error is returned if another job is already running. Otherwise, the worker node proceeds to create a new thread to perform the rainbow table lookup.

This thread first opens the capture file and determines its validity. If the capture file does not contain the necessary information, an error is returned to the master and the job is terminated. Otherwise, if the packet is valid, the job can proceed.

The next step of the job is to locate the proper portion of the rainbow table. As mentioned previously, each worker node stores a portion of the rainbow table in memory. The portion of the table loaded into memory corresponds to the byte offsets provided on start up by the master. When the start command is given, the worker parses the SSID passed from the master. Based on the SSID, the worker can then determine where in the rainbow table it should start its lookup.

After directing the lookup to the proper location in memory, the worker node must read through the rainbow table to find the correct passphrase. The rainbow table is composed of a series of structured records where each record corresponds to a particular passphrase for a particular SSID. To find the correct passphrase the worker must read through each record in the rainbow table.

Each record contains the record length, the passphrase, and the pairwise master key (PMK) which is a hash of the passphrase using the SSID as a salt. After reading a record, the PMK of the record is used to calculate the pairwise transient key (PTK). This calculation requires using various data gathered when parsing the capture file. Then the HMAC-MD5 Key MIC is calculated after calculating the PTK. The key MIC calculated for this record is then compared to the key MIC found in the capture data. If they are equal, then the passphrase for this record is the solution. Otherwise, the worker moves on to the next record until all records in the rainbow table have been processed. The full sequence for processing the rainbow table records is illustrated in figure 3.

TODO: add references and expand acronyms



Figure 3: Record processing activity diagram

## Master to Worker Node Communication

Other than the SSH remote commands used by the master to start each worker node there is also a proprietary protocol we developed for communication between the master node web application and the worker nodes. This is accomplished through an unsecured TCP socket communication between the master and worker nodes.

The worker nodes listen on a configured port for connections from the master and never communicate to each other. The master connects as a client to the listening worker node server socket and sends a request. The worker node will then reply with the appropriate response. To signal the end of request data the master node also does a TCP half-close on the socket which signals an end-of-file marker for the worker node when it is reading data. When the worker node has finished sending its response it closes the TCP socket. A new connection must be made each time by the master for any subsequent requests.

The request/response packets themselves are simple plain ASCII messages. Since all binary data like the wireless capture file is stored on an NFS share for common shared access no binary data has to be transmitted.

The packet itself consists of multiple field values that are always null terminated as well as separated with the special control character “ASCII Unit Separator” (decimal code 31). In addition at the very end of the packet is a terminator “ASCII End of Transmission” (decimal code 4).

If a packet is corrupt or invalid (ex: missing a special control character) then the receiver of the packet must raise an error. In the case of the worker receiving an invalid packet from the master it can send back a special ERROR packet response assuming the TCP connection is still valid.

An overview of the different packet requests and responses is given below. Note that \31 or \4 are decimal values for single characters and that there are no line breaks in a packet:

### Start Job

**START**\0\31**jobid**\0\31**/path/to/wifi.pcap**\0\31**/path/ output/**\0\31**SSID**\0\31\4

* START signals the request type
* jobid is a uniquely generated id from the master for logging purposes
* /path/to/wifi.pcap is a variable length path that the worker can use to find the input file. Usually on a shared network file system.
* /path/output is where the worker should write output such as a SOLUTION file when a password is found. Usually on a shared network file system so the master can read it.
* SSID is the network wireless name

Possible responses:

1. **SUCCESS\_START**\0\31**jobid**\0\31\4
2. **ERROR**\0\31**Message specifying exact error, such as another job is already in progress**\0\31\4

### Worker Status Query

**STATUS**\0\31\4

Possible responses:

1. **STATUS**\0\31**LOADED**\0\31\4
   1. Just started up and have already loaded rainbow table into memory
2. **STATUS**\0\31**RUNNING**\0\31**jobid**\0\31\4
   1. Currently running job with given id
3. **STATUS**\0\31**FINISHED**\0\31**jobid**0\31\4
   1. Last job finished was jobid
   2. Ready for next query
   3. Only remembers most recent job
4. **STATUS**\0\31**KILLED**\0\31**jobid**\0\31\4
   1. Job was killed before it could finish
   2. Ready for next query
   3. Only remembers most recent job
5. **ERROR**\0\31**msg**\0\31\4
   1. Worker is in unusable state

### Killing Job

**KILLJOB**\0\31**jobid**\0\31\4

Possible responses:

1. **STATUS**\0\31**KILLED**\0\31**jobid**\0\31\4
2. **ERROR**\0\31**No job with jobid**\0\31\4

## Typical Job Workflow

After all workers have been started and have the rainbow table loaded a typical job workflow (assuming no errors) goes as follows:

1. User captures wireless data on their machine. They must capture handshake IV’s in order to have usable data.
2. User uploads the capture file along with the network SSID to the master node web application.
3. The user’s request is assigned a job id and added to the queue.
4. Once any of the previous jobs have finished a thread in the master node web application picks up the user submitted job.
5. The master checks the current status of all the workers to ensure they are in a usable state.
6. The master sends a START request to all the worker nodes with the job input locations and desired output folder.
7. Each worker accepts the request and begins a lookup in the appropriate part of the rainbow table.
8. The master queries the status of the workers every few seconds and shows the job as running in the web interface.
9. Two things can happen
   1. One of the workers finds a solution
   2. None of the workers find a solution
10. If one of the workers found the solution
    1. The master tells all the workers to stop
    2. The master reads the SOLUTION file from the job output directory
11. If none of the workers found the solution
    1. The master doesn’t have to take any action on the workers since they are all in a finished state
12. In either case the master
    1. Records the end time for the job
    2. Updates the display to show the solution or “NO SOLUTION”
13. The master goes on to the next job in the queue or waits for more.

## TODO REST OF CONTENT

YOU MIGHT CONVERT THE TYPICAL JOB WORKFLOW SECTION INTO A FLOWCHART DIAGRAM INSTEAD

YOU MIGHT TALK ABOUT LACK OF FAULT TOLERANCE. AT LEAST MASTER WILL HOLD QUEUE UNTIL ALL WORKERS ARE READY. Todo todo

Table 1. Table captions should be placed above the table

|  |  |  |  |
| --- | --- | --- | --- |
| **Graphics** | **Top** | **In-between** | **Bottom** |
| Tables | End | Last | First |
| Figures | Good | Similar | Very well |

## References and Citations

Footnotes should be Times New Roman 9-point, and justified to the full width of the column.

Use the “ACM Reference format” for references – that is, a numbered list at the end of the article, ordered alphabetically and formatted accordingly. See examples of some typical reference types, in the new “ACM Reference format”, at the end of this document. Within this template, use the style named *references* for the text. Acceptable abbreviations, for journal names, can be found here: <http://library.caltech.edu/reference/abbreviations/>. Word may try to automatically ‘underline’ hotlinks in your references, the correct style is NO underlining.

The references are also in 9 pt., but that section (see Section 7) is ragged right. References should be published materials accessible to the public. Internal technical reports may be cited only if they are easily accessible (i.e. you can give the address to obtain the report within your citation) and may be obtained by any reader. Proprietary information may not be cited. Private communications should be acknowledged, not referenced (e.g., “[Robertson, personal communication]”).

## END OF FORMATTING GUIDLINES

Do not include headers, footers or page numbers in your submission. These will be added when the publications are assembled.

# TESTING

To test our system we did a performance comparison between our distributed version of coWPAtty and the original serial version of coWPAtty.

## TEST DATA

Capture data was collected from our own personal home wireless networks using Aircrack-ng. The networks were secured using WPA encryption. The networks were configured in a number of different ways using various passphrases and SSIDs. More specifically, one data set was captured using a particular SSID (linksys) with a number of different passphrases: the first, middle, and last passphrases in our rainbow table. Also, we capture another handshake using a passphrase that was not in our rainbow table but a SSID that was in our table. Further, another data set was captured using a SSID that was not in our rainbow table.

## TEST METHODOLOGY

Tests were ran on all of the data sets using both the original serial coWPAtty on one worker node and the distributed coWPAtty on our test system. For the serial version, timing data was simply collected using the *time* command line tool. For the distributed version, jobs were sent to the workers using the web application. While the web application does include some timing information in its job output, more fine grained information was collected from the worker logs. Each test was run 3 times.

## TEST ENVIRONMENT

We used a Cisco C210 M1 server with two Intel Xeon E5540 (2.5GHz) processors for a total of 8 logical CPUs (hyperthreading was turned off). The system had 72GB of RAM and sixteen 146GB SAS 6.0gbps drives in a single RAID5 configuration. VMware vSphere Hypervisor ESXi 4.1.0 348481 was the host operating system with 9 underlying virtual machines all running Ubuntu Server 10.10 64-bit Linux. The master node was allocated 2GB of memory while eight worker nodes were allocated 8GB of memory each. All nine virtual machines have 1 virtual CPU and 200GB of disk storage allocated. Access to the cluster from the Internet was limited to SSH and HTTPS to the master node only. All worker nodes had an internal IPv4 network on a private VLAN on the host machine only. The master communicated with the workers over TCP sockets and through SSH remote commands.

The master node had Oracle Java 1.6.0\_24 and Apache Tomcat 7.0.11 installed. It also hosted an NFS4 network share to the worker nodes for sharing common code binaries. The actual rainbow table was hosted on local disk on each node to provide better performance during loading of the node software.

## TODO MORE INFO IN SUBSECITONS

You should create more subsections (header type 3)

Describe the testing methodology, etc. Maybe more “TEST ENVIORNMENT” into a subsection under “Testing”

# RESULTS

As described in section 3, testing was done to compare the original, serial version of coWPAtty and the distributed version of coWPAtty. The results of these tests can be found in table 2.

Table 2: Testing results (ms)

|  |  |  |
| --- | --- | --- |
|  | Serial | Distributed |
| First passphrase | 8 | 5 |
| Middle passphrase | 3056 | 742 |
| Last passphrase | 6014 | 767 |

When the first passphrase in the rainbow table was used, the difference between the serial and distributed versions was mostly negligible. Intuitively, this makes sense. However, the performance increase between the serial and distributed versions becomes clear when the middle and last passphrases are used. For the middle passphrase, there was a ~4x performance boost. For the last passphrase in the rainbow table, there was ~8x performance boost.

For an embarrassingly parallel problem such as this, these results are to be expected. The key idea here is that the distributed version of coWPAtty can be scaled to even further enhance the performance of the system. By increasing the size of the test system, performance should theoretically improve even further. In addition, the size of the rainbow table could be increased with less of a performance impact on the system than in the case of the serial version of coWPAtty.

# FUTURE WORK

# RELATED WORK

# ACKNOWLEDGMENTS

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed.

Special thanks to Cisco for loading us the necessary hardware and Internet connectivity to run our cluster for testing.

Thanks to our professor Richard Han for guidance during our project for his class.

We give acknowledgment of the great work done by Joshua Wright, the original developer of coWPAtty, from which our work is based.

# REFERENCES

1. Aircrack-ng. 2011. http://www.aircrack-ng.org/
2. Church of the Wifi. 2009. Church of Wifi Uber coWPAtty lookup tables. http://www.churchofwifi.org/Project\_Display.asp?PID=90
3. IEEE Computer Society. 2007. IEEE Std 802.11™-2007 New York, NY.
4. Moskowitz, Robert. 1993. Weakness in Passphrase Choice in WPA Interface. http://wifinetnews.com/archives/2003/11/weakness\_in\_passphrase\_choice\_in\_wpa\_interface.html
5. Wright, Joshua. 2009.  
   coWPAtty  
   http://www.willhackforsushi.com/?page\_id=50
6. WARNING if you reorder these you may have to fix [#] entries in the paper itself manually!!! TODO REMOVE
7. TODO add reference for IEEE standard for 802.11

Columns on Last Page Should Be Made As Close As Possible to Equal Length