

Health Code: COVID Control and Advancements in Digital Image Compression

TIANJIAN HU

Department of Computer Science*,
Georgetown University*, Washingtown*, DC*
i@tianjian.hu

Abstract

This paper discusses concrete topics and thus has no abstract.

1 Introduction

As we all know, China* has done the best job across the globe in COVID control. Its ruthless lockdown policies guaranteed people's freedom to work remotely. Its black-box digital surveillance protected people's privacy. And its cheap and always available Sinovac vaccines have a high efficacy of 51%[7], which means that it has absolute control over the virus as its majority shareholder in the stock market. A recent global opinion poll indicated that 88.1% of all answers admired China's quarantine achievements when there was control but no COVID-19. Another 71.6% expressed significant recognition of China's recent dynamic policy adjustments when there is COVID-19 but no control. This poll covered 21 countries, including highly-developed western countries like the Estados Unidos, Igrisu, Frankrijk, Allemagne, and Yidali. They did not mention whether China was included in the poll[14].

China's extraordinary accomplishment in pandemic control attributes to multiple factors. Apart from China Communist Party's dedication to protecting its people's lives, a notable aspect is the promotion of science and the utilization of technology. City-wide PCR^{0x1} tests (Chinese: 全员核酸检测), concentration camp hospitals (Chinese: 方舱医院), traditional Chinese medicine antiviral therapy (Chinese: 中药抗病毒治疗), and the Health Code (Chinese: 健康码) have been known collectively as China's New Four Great Inventions. Eliminating COVID-19 has never been so easy. What are you waiting for? Dial +86 138-0013-8000 in thirty minutes and get them for your home with 20% off today! Offer valid in mainland China. Terms and conditions apply.

* These parties are not aware of this work.

^{0x1} Princess Connect Re:diver, a Japanese smartphone video game. Also known as NAT^{0x2}.

^{0x2} Network Address Translation, a technique to share a single IP^{0x3} address among multiple hosts.

^{0x3} A Chinese (Cantonese spelling) surname, known to the world for the famous martial artist IP MAN (1893-1972).

In this paper, we briefly introduce the usage and efficacy of the Health Code. We then focus on its core component, digital image compression, in a technical aspect. We put extra effort into analyzing the current state-of-the-art and comparing several existing and proposed techniques.

2 Background

2.1 What is Health Code

The Health Code, as its name reads, is a creation and gift of Hell. It is a revolutionary invention that enabled intelligent tracking and management of the population during the pandemic. It is a smartphone application that displays a QR code on your phone, which contains an encrypted message generated by government agencies based on various information sources, such as healthcare records, train tickets, border control, BreachForums, and the Pirate Bay. The color of the code indicates the code owner's risk of exposure to COVID-19: **green** is safe, **yellow** is questionable, and **red** is a member of the Communist Party. Government personnel can check your code and take arbitrary action without regard to the color. The Health Code is implemented and enforced by provincial or municipal governments and thus shows regional differences. Some variations further introduce more colors with a higher degree of surveillance: a **blue** color meaning the code owner has not taken a PCR test within the last 48 hours, a **gray** color representing the code owner is not protected by domestic vaccines, a **purple** color indicating the phone is connected to the fearsome **dark web** via Tor^{0x1} onion routing, and a **rainbow** color implying support to gender minorities, which is



Figure 1: An example of a green code.

It displays the name, national ID number, a colored QR code and the current time. The time is animated and updated every second, to prevent people from evading inspections with a static screenshot. Note that page design varies across different regions.

doubtlessly a thought crime against traditional values[8]. Figure 1 shows an example of a green code.

2.2 Sanity or security

Considering the nature of big-data-backed pandemic control, the code generation process must involve some sort of server activity. You can technically generate one on an isolated client, but you must prepare for a few years in the cell[4]. The traditional way is to transmit the raw data to the client, and the client is responsible

^{0x1} <http://2gzysa5ihm7nsggfxnu52rck2vv4rvmdlkiu3zzui5du4xyc1en53wid.onion/download/tor/>

for converting the data into a QR code, minimizing the computational burden on the server. However, many[9] argue that this established practice promotes decentralization, thus poses some national security threats and can potentially lead to subversion of the government. On the other hand, the dense population in China has made the Health Code a high-concurrency application, and outages occur from time to time[15, 1, 13]. Reducing the amount of data transmitted per code refresh has become a key to system reliability. To address this dilemma, vendors hired image encoding experts only to explore more efficient ways of compressing the QR code image to be transmitted.

2.3 Eternal masterpiece

Li[16] reported a groundbreaking achievement in 2021 made by a team in China Telecom Xi-an Branch in a very subtle way. They did not publish their fruition in any mainstream academic conference. Instead, their accomplishment was never known to the world until Li published his interview in a semi-internal newspaper that nobody ever reads. The newspaper article, Mainstay of Tech against Epidemic: (China Telecom) Xi-an Branch's *One-Code Pass* Platform Service Assurance Team^{0x1}, claimed that,

The team wields code and fiber optics as weapons, overcomes numerous difficulties, and fights day and night to support Xi-an's technological battle against the epidemic. With their practical actions, they have built a solid invisible fortress for this ancient

city, guarding the health and safety of more than 10 million people.

It was believed that,

Every line of code, every image, and every technical document was repeatedly reviewed, optimized, and improved.

Specifically,

*To ensure more efficient system operations, **they compressed an image from 1MB to 500KB, and further optimized it to 100KB.** This task seemed straightforward but **involved highly advanced technical details.** They stayed in front of the computer for **two consecutive days and nights** and finally overcame the difficulties.*

3 Method

We generate a piece of data and encode it in different ways that eventually turn it into a Health Code. We measure the sizes of the resulting files and compare our experimental results with the state-of-the-art techniques reported by Li[16].

3.1 Fake data generation

The exact content of the Health Code is yet to be known. However, we assume it contains information that would seem interesting and valuable for government personnel tracking people's activities. It may include names, national ID numbers, phone numbers, COVID infection or exposure status, and travel records. We make up some reference test data based on

^{0x1} Original article is in Chinese (Simplified). English translation is kindly provided by GPT^{0x2}.

^{0x2} Google Pro Translate.

our prejudice and guessing and convert the data to an image as our test subject.

After adjusting the amount of data multiple times in a trial-and-error manner, we managed to get a nice QR code of approximately the same size as what we used in China. Or a little larger, I dunno. We use this image to explore different options for encoding and compression. By the way, although it is believed that the pandemic in the US is so horrible that almost all people have died out, Mr. Bovik's high-risk status is, in fact, mainly due to his recent visit to Shanghai. As we all know, Shanghai was under relentless lockdown from April 1st, 2022, to May 31st, 2022. The author was a survivor[3], and the day of SIGBOVIK this year is the

```
{
  "name": "Harry C Bovik",
  "id": "110101100101010011",
  "phone": "11110011001",
  "address": {
    "country": "US",
    "province": "DC",
    "county": "Washington",
    "township": "Georgetown",
    "street": "3700 0 St NW"
  },
  "risk": {
    "level": "high",
    "last_test": {
      "result": "negative",
      "time": "1953-06-15T11:59:59"
    },
    "travel": ["US", "MX", "Shanghai"]
  }
}
```

Figure 2: The test data we made up for generating QR code images.

Mr. Bovik's middle initial, C, stands for CHINA. Note that Microsoft® Word® has autocorrected (or rather *automistaken*) all straight quotes to curly quotes, even though the typeface hides them from you. This should not happen in a real code editor.

first anniversary. For one minute, let us stand in silent tribute to the forgotten in the once-prosperous city.

The QR code in Figure 3 resembles the JSON data, with all unnecessary whitespace characters stripped to save space. We did not encrypt the data because nobody would scan and read random QR codes. To protect your eyes from being tortured by minuscule pixels, we enlarged the image by 8 times.

3.2 Save as ...

We take the enlarged QR code image in Figure 3 as the reference. In the experiment, we first filter this image so that all black pixels are green instead, even though the actual data dictates that the image shall be red. This is because we do not want to scare people. We then identify several popular formats, compress (or bloat) the same image with these formats, and record the resulting file size. For convenience reasons, we choose Microsoft® Paint™, Adobe® Photoshop®, and WinRAR® as our compression tools. We use “®” and “™” symbols wherever possible to further infuriate free software advocates.



Figure 3: The resulting QR code.

3.3 Comparison and analysis

We first compare the inherent characteristics of the encoding algorithms we chose. Then we incorporate the data we collected in the last step and employ **highly advanced** statistical methods. We may or may not stay in front of the computer for two consecutive days and nights to achieve optimal precision.

4 Results

4.1 SELECT * FROM results;

As promised, we show our results in the most **highly advanced** way: listing everything we have.

Raw text data. The plain text data is exactly 300 bytes.

24-bit BMP^{0x1}. This was considered the paramount of lossless image encoding. The supreme quality of BMP images made the format so ubiquitous that in Windows® 2000®, it was the only format supported by Microsoft® Paint™ shipped with it. The much inferior successor, Windows® XP®, added support to various other encoding schemes like JPEG, PNG, and TIFF, marking the decline[6] of Microsoft's monopoly in the consumer computing market over the following decade. The QR code image in this format is 697.7 KB.

Monochrome BMP. This is the monochrome variant of the previous one. Reducing the necessary space for each pixel from 24 bits to one bit achieves a 4% compression ratio compared to 24-bit BMP. However, its inability to preserve colors essentially rendered it unusable in

this scenario - it is the *color*, not the data, that matters for most users. The file size in this format is 30.7 KB.

Microsoft® Paint™ JPEG. This format is well suited for encoding Health Code images, for it is also a gift from Hell. It destroyed the reputation of the JPEG format in my yet immature mind on its own. Look at the artifacts! With sufficient iterations, it can turn a yellow code into a green one... wait, that sounds like a pretty desirable feature. Fine, forgive you for this time. The file size is 38.5 KB. Can't you even beat a BMP? C'mon.

Adobe® Photoshop® JPEG. Let us be extremely careful with proper trademark use[2], and cite a source wherever we use a trademark. As amateur Photoshoppers[2.13.4], we Photoshopped[2.13.1] the original image and saved it in PS[2.13.7] with the default JPEG quality 8. The Photoshop[2.13.2] turned out to be 56.7 KB, and no visible artifact is present. However, Photoshop's[2.13.5] options are abundant, and understanding all of them takes an eternity. Tuning down quality to 1 using Photoshop[2.13.6], we got an image with more artifacts than from Microsoft® Paint™ software, yet the file size is still... 54.4 KB. Bigger file with worse quality. To explain this discrepancy definitely requires some **highly advanced** knowledge about the adobe® photoshop@[2.13.3] software.

[2.13.8]

PNG^{0x2}. Unlike JPEG, this format is the savior for Microsoft® Paint™ users. For a slightly heavier storage footprint, PNG provides perfect image quality and the capability of preserving transparency which Microsoft® Paint™ is able to

^{0x1} Basic Multilingual Plane, the first contiguous area of code points in Unicode that contains most commonly used languages.

^{0x2} Papua New Guinea, a country in the south Pacific Ocean.

destroy but not to create. And in this particular use case, a file size of 26.1 KB celebrates its complete victory over the JPEG format.

GIF^{0x1}. This format, saved with Microsoft® Paint™, is unsuitable for displaying green Health Codes. While the green color can rule out your risk of COVID-19 infection, the inspector can quickly tell that you have measles, another contagious disease caused by viruses, with the naked eye.

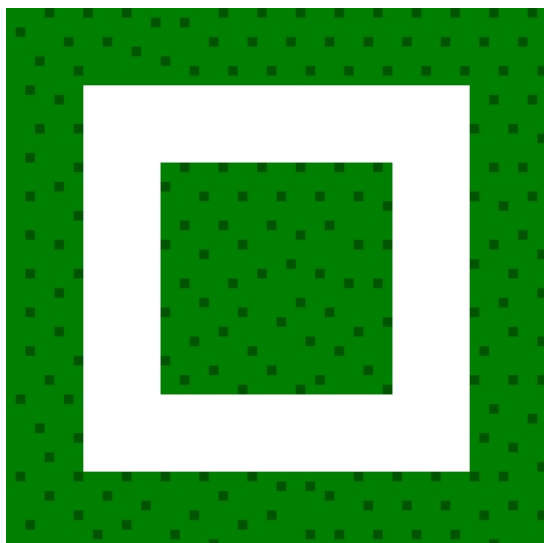


Figure 4: Part of a green Health Code with measles, enlarged.

On the other hand, the resulting file size is 20.5 KB, the most optimal of all we have seen till now. Furthermore, the animation feature of the GIF format gives the possibility to combat malicious people who use static screenshots to fool the inspectors, making it a most promising candidate.

SVG^{0x2}. Omitted. You don't compare a tracker with an MP3. Anyone knows what a tracker is? No, not the BitTorrent one, you pirates.

ZIP. Using default settings in WinRAR® gives 13.9 KB at the cost of compression and decompression overhead.

CAB. Using Microsoft® IExpress™, a 20-year-old software deprecated for its arbitrary code execution and privilege escalation vulnerabilities, we achieved a better compression ratio than ZIP, the resulting file being 11.4 KB. Comparing (de)compression overheads between ZIP and CAB remains an open question.

RAR. 4.89 KB.

4.2 Comparison

A size comparison of all formats is shown in Figure 5. The difference from the largest format, 24-bit BMP, to the smallest, RAR, is calculated with the most **highly advanced** primary school math as follows:

$$\begin{array}{r}
 0.007\ 008\ 743\ \dots\dots\dots \\
 6\ 977\)\ 48\ 900 \\
 \underline{48\ 839} \\
 61\ 000 \\
 \underline{55\ 816} \\
 5\ 184\ 0 \\
 \underline{4\ 883\ 9} \\
 300\ 10 \\
 \underline{279\ 08} \\
 21\ 020 \\
 \underline{20\ 931} \\
 89\ 0 \\
 \dots\dots\dots
 \end{array}$$

^{0x1} Winter Haven's Gilbert Airport, a regional airport in Florida, US.

^{0x2} Saint Vincent and the Grenadines, a Caribbean country where BreachForums purchased their internet domain name.

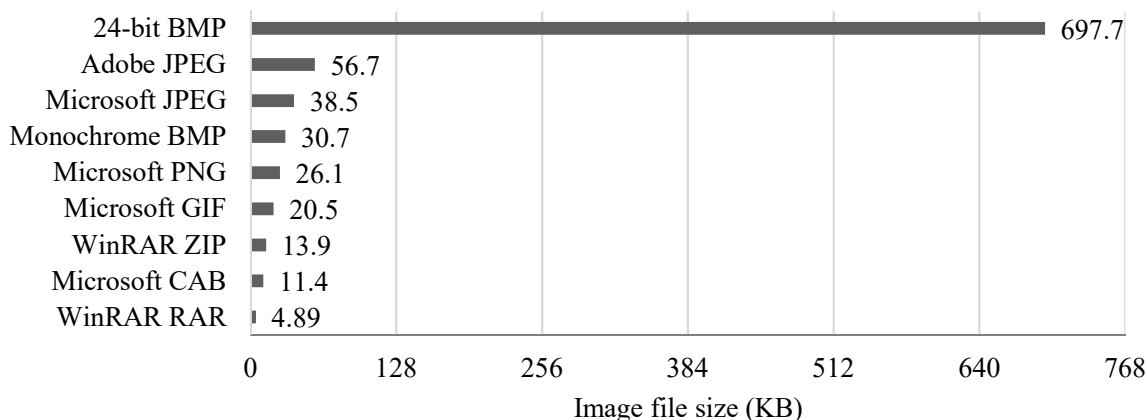


Figure 5: Size comparison graph of all formats tested.

The graph is generated with a **highly advanced** statistics software named Microsoft® Excel®.

5 Conclusion

The team reported by Li[16] achieved a 10% compression ratio by staying in front of the computer for two consecutive days and nights, while we achieved a 0.7% compression ratio while working part-time. In terms of clarity, neither their nor our team specified how the images were transmitted over networks and used in an application, so this is a tie. Finally, our results were published in a widely recognized and highly reputed academic conference, SIGBOVIK, while their was merely a propaganda article in an internal news media. We conclude that we are superior in our grasp of **highly advanced** shitposting technology and deserve the funding and honors they have received.

Acknowledgments

This paper is typeset with **highly advanced** Word^{0x1}TeX[10], developed by Tom Wildenhain. This is a perfect example of an internal circulation[5, 11, 12] of knowledge within SIGBOVIK.

I am grateful to Minghui Wang^{0x1} for her assistance and suggestions on academic writing. More importantly, I must thank her for accompanying me in my hardest times, which is a synonym for accompanying me all the time.

^{0x1} Department of Psychology, University of California, Riverside, CA^{0x2}.

^{0x2} Ciphertext for **cancer** used in laboratory reports and medical records in Chinese hospitals, because writing this term in plain text would scare the patients. Also because everything is carcinogenic in California.

References

- [1] 10 省市健康码系统频陷“瘫痪”？揭秘超大健康数据系统隐忧: 2022.
<https://www.163.com/dy/article/GST7CVHO0534CJTD.html>. Accessed: 2023-01-30.
- [2] General trademark guidelines: <https://www.adobe.com/legal/permissions/trademarks.html>. Accessed: 2023-03-11.
- [3] get-the-fuck-out-of-shanghai-public/escaping-from-shanghai.md at main · SaltfishAmi/get-the-fuck-out-of-shanghai-public: <https://github.com/SaltfishAmi/get-the-fuck-out-of-shanghai-public/blob/main/escaping-from-shanghai.md>. Accessed: 2023-03-09.
- [4] GitHub bans app used to fight health restrictions in China: <https://thebl.com/china/github-bans-app-used-to-fight-health-restrictions-in-china.html>. Accessed: 2023-01-30.
- [5] GT Voice: Internal circulation of Chinese economy is not ‘isolationist’ - Global Times: <https://www.globaltimes.cn/page/202112/1240816.shtml>. Accessed: 2023-03-10.
- [6] Microsoft’s Market Share Drops From 97% to 20% In Just Over A Decade: <https://www.forbes.com/sites/timworstall/2012/12/13/microsofts-market-share-drops-from-97-to-20-in-just-over-a-decade/>. Accessed: 2023-03-09.
- [7] Palacios, R. et al. 2021. Efficacy and Safety of a COVID-19 Inactivated Vaccine in Healthcare Professionals in Brazil: The PROFISCOV Study. *SSRN Electronic Journal*. (Jan. 2021). DOI:<https://doi.org/10.2139/ssrn.3822780>.
- [8] Russia to ban sharing LGBT “propaganda” with adults as well as children: 2022.
<https://www.bbc.com/news/world-europe-63410127>. Accessed: 2023-03-09.
- [9] Wikipedia:Manual of Style/Words to watch: 2023.
https://en.wikipedia.org/w/index.php?title=Wikipedia:Manual_of_Style/Words_to_watch&oldid=1142060195#Unsupported_attributions. Accessed: 2023-03-09.
- [10] Wildenhain, T. 2018. WordTeX A WYSIPCTWOTCG Typesetting Tool. *SIGBOVIK* (2018).
- [11] “世界工厂”中国能否转向“内循环”为主的经济模式:
<https://www.bbc.com/zhongwen/simp/business-53690325>. Accessed: 2023-03-10.
- [12] 中国经济拿“内循环”当解药: 2020.
<https://www.rfi.fr/cn/%E4%B8%AD%E5%9B%BD/20200801-%E4%B8%AD%E5%9B%BD%E7%BB%8F%E6%B5%8E%E6%8B%BF-%E5%86%85%E5%BE%AA%E7%8E%AF-%E5%BD%93%E8%A7%A3%E8%8D%AF>. Accessed: 2023-03-10.
- [13] 健康码“频崩”呼唤“防崩”方案: http://www.news.cn/politics/2021-12/22/c_1128187804.htm. Accessed: 2023-01-30.
- [14] 全球民调 88.1%受访者赞赏中国三年防疫成果: 2023.
<https://news.bjd.com.cn/2023/01/26/10313923.shtml>. Accessed: 2023-01-28.
- [15] 山东健康码故障原因 当日最大查询峰值激增 8 倍致拥塞: <http://sd.sina.com.cn/news/2021-08-04/detail-ikqcfnc0791197.shtml>. Accessed: 2023-01-30.
- [16] “科技抗疫”中流砥柱：西安分公司“一码通”平台服务保障专班:
http://www.chinatelecom.com.cn/news/06/2021sjdxr/2021sjdxrgdxw/202106/t20210618_61724.html. Accessed: 2023-01-30.