A NetLogo plug-in to secure data using GNUs Pretty Good Privacy software suite

Doug Salt Gary Polhill

Abstract

A description of a NetLogo plugin and the reasoning behind its design and implmentation. The plugin makes use of Gnu's Pretty Good Privacy software suite to encrypt arbitary data sources in Netlogo. This both secures the data to a reasonable degree and protects any sensitive data that might be in use for a publically available model.

Introduction

Cheap, publicly-accessible, distributed storage, colloquially known as the "cloud" is becoming increasingly prevalent [@] and is increasingly used for the storing of experimental data [@]. Storing experimental data in this way has several advantages, such as any access to the internet allows instant access to this data [@]. This means the data is effectively accessible anywhere. Data might be stored in a single location, or it might be distributed. The convenience is that from the point of view of the consumer of the data, it all appears to originate at a single web-based location [@]. Also such data is cloned and distributed in physical space for the purposes of fault-tolerance [@] and thus can exist in many locations simultaneously [@]. Thus the chances of it being lost are remote [@].

It is reasonably easy to use such cloud hosted data in NetLogo models. Some institutions provide their own, cloud-based solutions, but most researchers will use at least one of the following, major cloud storage providers such as Dropbox [@], Microsoft's OneDrive [@] and Google Drive [@]. This list is by no means meant to be comprehensive. It is our suspicion that some, less technically aware NetLogo users are making use of the cloud without being aware of the ramifications of doing so, but doing so because these resources are extremely convenient, cheap or more often free. This problem may well arise because each of these hosting companies provide tools that allow user-transparent, local mounting of such resources. These exist for the most prevalent platforms such as all Microsoft Windows versions greater than 7 [@]; all versions of Android [@], and Apple's two operating systems: OSX and IOS[@]. This means that the "cloud" storage appears as local storage on the local machine, and NetLogo models do not need to be changed to access such data (other than changing a file name). Indeed some users may not even be aware that the data they use is in the "cloud" already.

The advantages listed above are also the method's disadvantages. The publicly accessible nature of such data could violate regional privacy laws. For instance storing personally identifiable data of a sensitive nature without sufficient safe-guards now violates the European GDPR [@]. The multiplicity of the storage means the creator of the data has largely lost control over the destruction of the data. Most users of cloud data are unaware that even their "scratch" data is stored in the cloud. That is intermediate files or snapshots of works-in-progress. This becomes problematic when regulations or ethics require that data is permanently and effectively deleted. Also if such data is of a personally identifiable nature, then, given GDPR requirements it is a legal requirement of the researcher to store the data in particular geographical areas and moreover ensure that when usage conditions specify deletion, then deletion must, absolutely have taken

place (ibid.).

The *only* way to ensure that effective deletion takes place when using such utility computing infrastructure is to encrypt the data sufficiently that when key for decrypting such data is withheld then the original data can no longer be retrieved [@]. This is reasonably easy to achieve given that, with current technology a brute-force attack on a 128 bit AES encoded data would take on average $1.02 \times 10_{18}$ years to work. (https://www.eetimes.com/document.asp ?doc_id=1279619). Doubling the size of this key to 256 bits is thought to effectively protect such data from proposed attacks such as those theoretically available if quantum computing proves to be successful [@]. Destroying or withdrawing the encryption key therefore effectively deletes such data. Thus encrypting data has both the desirable properties of securing and ensuring appropriate deletion of the data.

It should be noted that most "cloud" provision does, as standard practice, encrypt users' data [@]. The problem is that the provisioning entity controls the keys, and is is not entirely clear what jurisdictional laws to apply given the international nature of such providers. For example some doubt over data jurisdiction currently exists between the European Union and the US government (http://ec.europa.eu/justice/data-protection/international-transfers/adequacy/index_en.htm). Thus it is impossible for a user of such services to guarantee the correct jurisdictional standards are applied to their data, unless they take control of the encryption themselves.

Having established the need for encryption, the remainder of this paper will describe the installation and usage of NetLogo extension that will allow the easy decryption of previously encrypted data sets. This extension will require the installation of GNU's Pretty Good Privacy suite of programs, or at the very list have the command `gpg` in the execution path currently invoking the NetLogo model. This will be invoked in the background in order provide asymmetric and symmetric decryption for any data sets. Each of the possible use-cases will be described that the extension has been designed to address by way of a small example of usage. This will be followed by the usual discussion of issues raised by the utility and use of this plug-in.

The NetLogo Extension

Given that there is a need for such an encryption utility the problem becomes how can such user-controlled encryption be implemented in a user-friendly manner with minimal development. The last condition is important because coding encryption correctly is a hard problem [@]. Insufficient expertise can lead to attack opportunities due to weakness inherent in the developers' approaches [@]. It therefore makes a great deal of sense to use existing and proven software. In addition there is a requirement that users be able to encrypt their data in the first place. This rules out the usual practice of utilising an existing, programmatic libraries, created for specifically for the purposes of encryption/decryption. Such libraries are indeed proven, but usually lack the user-friendly encryption tools required to do the initial encryption. Such tools although usually trivial to create, crucially, still have to be developed and moreover, documented. Such requirements contain the possibility of the introduction of bugs. Additionally the use of such libraries requires the constant updating of the plug-in software, each time the library is updated - say due to the discovery of a new attack or bug. Such constraints can be mitigated by the use of an external software suite. That is, a NetLogo extension can be designed in such a manner to make calls to an "external" program. An external program in this context is software that is independently installed on a computer, is independent of NetLogo, and does not require NetLogo to work. An example of this approach is the NetLogo R extension [@] which obviously requires the independent installation of the R programming suite for it to work with NetLogo. Thus, if any problems are found with the external program, then just the external program needs updating. This does have the disadvantage of introducing an additional step in the utilization of NetLogo, but this is balanced not only be the addi-

tional utility and possible multiple uses of the external software suite, but by the huge reduction in complexity required to create the NetLogo extension. This has benefits in terms increasing stability and formal correctness for the extension.

The external tool chosen is GNU Privacy Guard, hereafter referred to as GPG. This is a well known suite of programs that at its heart uses OpenPGP standard as defined by RFC4880 (also known as PGP) [@]. Although designed primarily for the purposes of safeguarding communications, GPG allows the encryption of data; it features a versatile key management system, along with access modules for all kinds of public key directories. GPG is a command line tool with features for easy integration with other applications. The software is mature in that it was created in 1996 [@] and is widely used [@]. GPG provides a series of command line tools and is available on virtually every single computing platform. The presumption will be that GPG has been installed on the platform that is to run the NetLogo extension.

Because the extension uses GPG, the extension is very small and requires the installation of just one jar file. The extension is written in Scala [@] and built using sbt[@] and consists of the following primitives:

- `gpg:cmd`
- `gpg:home`
- `gpg:open`
- `gpg:read-line`
- `gpg:at-end?`
- `gpg:close`

The normal flow would look like that shown in fig. 1. We have tried to keep the operational semantics as natural and terse as we can make them.

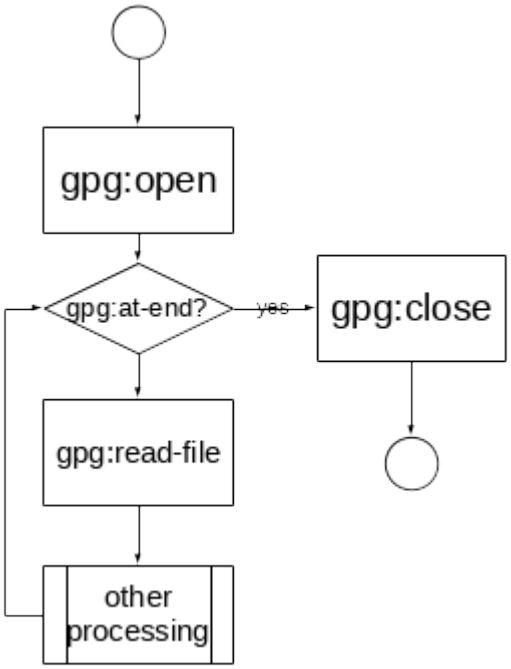

Figure 1: Typical extension flow

The installation jar can be found at https://gitlab.com:doug.salt/gpg.git. The file

`target/scala-2.12/gpg.jar`

should be copied to a file named gpg.jar. This file should be placed in the extensions directory of the NetLogo installation. This is normally:

- On Mac OS X: `/Applications/NetLogo 6.0.4/extensions`

- On 64-bit Windows with 64-bit NetLogo or 32-bit Windows with 32-bit NetLogo: `C:\Program Files\NetLogo 6.0.4\app\extensions`
- On 64-bit Windows with 32-bit NetLogo: `C:\Program Files (x86)\NetLogo 6.0.4\app\extensions`
- On Linux, or other *nix: the app/extensions subdirectory of the NetLogo directory extracted from the installation .tgz

Or, alternatively it can be placed in a sub-directory with the same name as the extension in the same directory as the source for the NetLogo model if the extension is not to be used globally. So for instance, this extension is known as `gpg` so if the model `example.nlogo` was placed in the directory `/data/models` the extension would have the path `/data/models/gpg/gpg.jar`.

The extension is invoked in the NetLogo code by adding the keyword `gpg` to the extensions keyword beginning the NetLogo model code. For example in the code examples provided for the gpg implementation, the following appears at the top of the code section indicating that the `gpg` and `csv` extensions are to be used.

```
extensions [gpg csv]
```

A brief description of the available extension keywords now follows.

gpg:cmd

This sets the path of the `gpg` command if the `gpg` command is not in `$PATH` for *nix system or `%PATH%` for Windows based systems. Its also allows the specification of additional parameters to `gpg`. This should not be needed. The only parameters that should require changing are the home directory containing the keyring. However, this can also be done using `gpg:home`. This multiple way of achieving the same end is due to OS sensitivity over paths. `gpg:home` provides an operating system agnostic method of specifying the keyring directory. `gpg:cmd` also allows the GPG executable to be wrapped, or replaced with something else. This may appear to be a security weakness, and indeed it is, but the choice of calling the program externally to NetLogo implies that this can be done without `gpg:cmd`. That is, it is easy to replace the `gpg` binary with something nefarious. This also applies to any non-static library that NetLogo makes use of. So, although not air-tight security, this does offer "reasonable" security. The only way to obviate such a weakness would be to statically compile in such libraries and this has consequences for security and flexibility as elaborated earlier.

Some examples of the invocation of this command might be

```
gpg:cmd "/opt/gpg/bin/gpg"
```

or

```
gpg:cmd "gpg --homedir ~/some-directory"
```

Note, the state of execution string will persist, and not reset until the next invocation of `gpg:cmd`. The command can be cleared to default by using either

`gpg:cmd ""` or `(gpg:cmd)`

gpg:home

This sets the home directory relative to the directory in which the NetLogo model resides. If this command is not used then GPG assumes the that its key ring resides in the sub-directory `.gnupg` of the standard home directory for that system.

Examples of the usage of this command might be

```
gpg:home .keyrings
```

This would expect the keyrings to be found in a directory `.keyrings` immediately below the directory in which the NetLogo code for the model resides.

Note, the home-directory will persist, and not reset until the next invocation of `gpg:home`. The command can be cleared to default by using either

`gpg:home ""` or `(gpg:home)`

The home directory can also alternatively be set using:

```
gpg:cmd "gpg --homedir ~/some-directory"
```

gpg:open

This attaches and decrypts a given cryptogram.

If `cryptogram_path` is the filename of the cryptogram and `cryptogram_id` will be the variable holding the id of the attached and opened cryptogram, and in addition the cryptogram has not been encrypted symmetrically, nor has does the key that has encrypted it require a passphrase, then this command would be used in the following manner.

```
let cryptogram_id gpg:open cryptogram_path
```

If the cryptogram `cryptogram_path` has been symmetrically encoded, or the its decoding key requires a passphrase then this can be specificed in the following manner, where "some-passphrase" is the required phrase.

```
let cryptogram_id  (gpg:open
    cryptogram_path "some-passphrase")
```

This will exception if the `cryptogram_path` does not exist, cannot be opened, or requires a passphrase when none has been supplied..

gpg:read-line

Reads a line of clear text. from a previously opened cryptogram. The file must have been successfully opened using `gpg:open`.

If `clear-text` is a previously declared NetLogo variable and `cryptogram_id` is the variable holding the id of the attached and opened cryptogram, then this command would be used in the following manner.

```
set clear-text gpg:read-line cryptogram_id
```

This will exception if the `cryptogram_id` does not represent a cryptogram that is attached and opened.

gpg:at-end?

Tests whether there are additional lines of plain text available for `gpg:read-line` to obtain. The file must have been successfully opened using `gpg:open`.

If `cryptogram_id` is the variable holding the id of the attached and opened cryptogram, then this command would be used in the following manner.

```
if gpg:at-end? cryptogram_id [
    ...
]
```

This will exception if the `cryptogram_id` does not represent a cryptogram that is attached and opened.

gpg:close

Closes and detaches the cryptogram. The file must have been successfully opened using `gpg:open`. This means the data is no longer sitting in memory encrypted.

If `cryptogram_id` is the variable holding the id of the attached and opened cryptogram, then this command would be used in the following manner.

```
gpg:close cryptogram_id
```

This will exception if the `cryptogram_id` does not represent a cryptogram that is attached and opened.

Illustrations

Symmetric encryption

This is the easier kind of encryption to understand. A secret is encrypted with a key to produce a cryptogram. That key, and cryptogram are then passed to the person who wishes to decrypt it. This person then uses the key to decrypt the cryptogram in order to obtain the secret. Until the advent of asymmetric, this was most usual method of encryption usage. The weakness with this approach is that the key needs to be transported along with the cryptogram.

Say we have some clear text containing some sensitive data, and we wish to encrypt this a key, say the string, "some-string", then the procedure using GPG would be the following:

```
gpg --encrypt \
  --passphrase "some-string" \
  --output cryptogram.gpg \
  clear.txt
```

The passphrase "some-password" and the cryptogram, `cryptogram.gpg` are now passed to the recipient who wishes to make use of the data in their model. Possessing both these components, then the code required in order to decrypt and make use of such data would be as follows:

```
  let file (gpg:open cryptogram
    "some-string")
  while [ not (gpg:at-end? file) ] [
    output-show gpg:read-line file
  ]
  gpg:close file
```

This is not particular secure as the key is embedded in the code. This insecurity could be reduced by requiring the pass-phrase to supplied in an automatically clearing input field in the NetLogo interface.

The primary advantage of this approach is its simplicity and it obvious semantics which make it easier to follow than the asymmetric key approach in the next section.

Asymmetric encryption

This is the most powerful facility of GPG. Asymmetric encryption offers the ability for any individual to encrypt a message, but only specific individuals being able to decrypt the file, *without having passed any encryption keys*. This is achieved by encrypting the file with the public key of the recipient, so only the private key of the recipient can then unencrypt the cryptogram. This makes this form of encryption enormously secure, and hard to exploit, because the key is never exposed to other parties. This is the unique appeal of key asymmetry: the only people who can open the file must be in physical possession of the private key, and if a passphrase is used, then they must also know something as well.

Asymmetric key encryption tends to confuse people [@]. It may, however, be thought of in the following manner. Consider a chest which has two locks on it. The first lock is a deadlock and may only be locked permanently with a key, otherwise that lock is always open. If this lock is locked, then this triggers the latching of a second lock. The first key corresponds to the public key, the second to the private key. In this scenario, if a secret is locked in the box, by the public key, this causes the second lock to latch and lock. The box may only be opened if and only if we have both the public and private key. This is not quite how asymmetric encryption in GPG works, but is near enough to give a reasonable understanding of the principles and its implications. For instance using this box system we can pass a secret to a person who owns the private key, safe in the knowledge that once this box is locked only they can unlock it. GPG is effectively just a method of leaving many copies of such boxes and many copies of such locking, public keys just lying around, just waiting to be used.

The code below presumes a cryptogram with no passphrase on the private key. So if we have some user, denoted `aUser`, and this user has a public key `aUser.pub`, an email associated with this public key of `aUser@anInstitution.ac.uk`, a private key, `aUser.ppk` corresponding to the public key `aUser.pub` and the clear-text in `clear.txt` is available in the same directory containing the NetLogo model and code.

Firstly the public key would need to be imported into the keyring of the person performing the encryption:

```
gpg --import aUser.pub
```

This user would then encrypt the clear text in `clear.txt` using the following command:

```
gpg --encrypt \
  --output cryptogram.gpg \
  --recipient aUser@anInstitution.ac.uk \
  clear.txt
```

The cryptogram is now encoded in the file `cryptogram.gpg`, with the public key, `aUser.pub`. To be able to decrypt the cryptogram, `cryptogram.gpg`, then the user who wishes to do the decryption must have the private key, `aUser.ppk` in their keyring. This may have happened in only two ways. Firstly `aUser.pub` the public key was generated by the command:

```
gpg --gen-key
```

This generates a private key into a person's keyring and moreover associates that private key with a particular email address. To obtain the public key then the following must be run:

```
gpg --export  aUser@anInstitution.ac.uk \
  > aUser.pub
```

This is the public key and may be distributed to anybody. There are no privacy implication on the distribution of this key.

Or, alternatively they would have had to import the private key into the keyring, say something along the lines of

```
gpg --allow-secret-key-import aUser.ppk
```

Given all the above, if the file `cryptogram.pgp` is present in the same directory as the NetLogo code and model. Also given the preconditions above, then the code to decrypt and show the code would be the following:

```
let file (gpg:open "cryptogram.gpg")
while [ not (gpg:at-end? file) ] [
  output-show gpg:read-line file
]
gpg:close file
```

This means that only a user in possession of `aUser.ppk` can decode the cryptogram `cryptogram.pgp`. Moreover if there is passphrase associated with `aUser.ppk` then this must also be supplied, further reducing the possibility of the cryptogram becoming compromised. The passphrase should really be supplied by an automatically clearing field provided in the interface. However we cannot enforce it, but only recommend this as the implementation.

Even better is that multiple email addresses, corresponding to multiple public keys can be provided at the point of encryption. This means the recipients can be limited to a specific set of individuals if required, and only those individuals can decrypt the cryptogram.

Reading an asymmetrically encoded CSV file

With a simalar set-up as above; then if we have some user, denoted `aUser`, and this user has a public key `aUser.pub`, an email associated with this public key of `aUser@anInstitution.ac.uk`, a private key, `aUser.ppk` corresponding to the public key `aUser.pub`. Also a clear-text comma separated variables file, in `clear.csv` has been enecrypted using `aUser.pub` to produce a file `cryptogram.gpg` in the directory that the NetLogo model and code resides. Then to decode the file we would use the following code.

```
let file gpg:open "cryptogram.gpg"
while [ not (gpg:at-end? file) ] [
  output-show (csv:from-row
      gpg:read-line file)
]
gpg:close file
```

This example is very similar to that of the previous section, and all it shows is that the extension can be used when coupled to other extensions available in NetLogo thus increasing its possible utility.

Discussion and conclusions

The real goal here is to encourage the use of asymmetric encryption, for the purposes of securing potentially sensitive data. The use of such encryption allows the personalisation of data in an unprecedented manner. That is, secrets can be passed between sender and

receiver based only on something publicly shared, something only the receiver has, and optionally knows (if there is a passphrase on their private key). What makes this method particularly secure is no exchange of key information has to take place. With the ability to encrypt for multiple recipients using a single cryptogram containing data, then this effectively means that encrypted data can be personalised to the group of receivers only: each having their own particular private key. This restricts the use of the data to that group, and that group only (providing their private keys remain uncompromised).

Such encryption has distinct advantages outside of this particular application, most notably for email. Consequently the infrastructure has been developed, and is in place to allow the publishing of public keys to the Internet, and their subsequent use in mostly email clients. Most notably there are a number of public key servers, where public keys may be retrieved on the basis of the supplied email address. However, the extra complexity to make use of this technology seems to have stalled its adoption. Additionally many of the free messaging service providers, such as Google and Microsoft have no interest in processing data which they can no longer read the content of, so have no incentive to develop easy-to-use implementations of asymmetric key encryption. To use this infrastructure to allow decryption and encryption of data in NetLogo, then this does have the extra dependencies of creating a private key; storing it (reliably and in a recoverable manner) and publishing the public key on a public key server. Public keys servers do have the problem of becoming cluttered with keys that are no-longer used, or many duplicate keys for the same email. It is probably best practice to make all keys time-limited for this reason. Such a facility is provided at the time of creation of the public key. However setting up a private key only needs to be done once. It can be done repeatedly if the problem of duplicates keys can be obviated. Indeed time-limited keys might be used as standard operational procedure, which would effectively mean the data is deleted upon expiration of the private key.

Setting up such a keys ensures that the data is secured against everything except code modification and memory sniffing. This reduces the security requirement footprint to a level where it might be considered that reasonable precautions had been undertaken to secure such data, or certainly as far as local legislation might require.

One of the central pressumptions of this paper is that the user of the data, whether this be a group or individuals, should be more responsible for the security of the data they use. This is enforced by the use of asymmetric encryption, forcing them to be in charge of their private keys in order that only the sender and the receivers are able to view and use the data. However, as mentioned this introduces a degree of technical expertise and knowledge may be beyond the average user, so another compromise might be the use of organizational key infrastructure. This would remove the onus of securing, and/or recovering keys from the user, but would represent a further compromise in the security of the data, in that anybody with sufficient permissions within a given organization would have unfettered access to the data. It might be argued that this is also reasonable security, and certainly a level of security that many organizations already adopt.

In conjunction with Infrastructure as a Service (IaaS), then it is becoming increasingly common to see NetLogo models and use NetLogo models in the cloud. The problem with cloud infrastructure is that it is non-localised and duplicated for the sake of fault-tolerance. It is therefore impossible to perform timely deletion of data in cloud infrastructure, because not even the service providers themselves will know the location and the number of copies of a given set of infrastructure at any one time. Key deletion is functionally and realistically equivalent to deleting the remote data. Encrypting data in such infrastructure, which may be vulnerable to bad actors via the service provider or the consumers of the infrastructure, should therefore standard practice to ensure that the data remains secure no matter what happens.

We have developed a plugin that uses the Gnu PGP software to allow various types of encryption on the data only. We could develop a plugin that obfuscates the code, but we believe that this not only violates

the code of openness that surround the NetLogo community, but also possibly violates the GNU Public License version 2 under which NetLogo is distributed. Taking somebody's open code and concealing it legally violates the license, as this is precisely the reason the license was created in the first place [@]. It also violates the principle of open science as people should be able to inspect models to see the reasoning that underlies them. This is increasingly important where such models are used for policy decisions [@]

This code has been tested on Linux, Windows 7, Windows 10, and OSX so far the code could be ported in entirety into NetLogo. There are java libraries available that mirror the functionality of PGP [@]. However this has the limitation of precluding the rapid release cycle of encryption software once vulnerabilities have been discovered. The code, as it stands is specifically designed for GPG, and the arguments expected by GPG. It would not be difficult to write wrapper scripts for other encryption suites, and therefore reuse this code. This, of course is a security weakness and does provide an additional attack surface, whereby the code could be replaced by compromised versions of the encryption suite, but as explained earlier, this provides a "reasonable" level of security.

This package is susceptible to memory sniffing attacks, particularly memory freezing attacks (crashing the application or entire machine such as those found in privilege-raising attacks[@])). These can be mitigated by encrypting disk swap, but if the key is in the clear anywhere in memory, then there is always a chance that it can be obtained. This will always be a weakness of any encryption system that is computerised[1].

As mentioned this is still too complicated for non-technical users. However those users that deal with sensitive information have a duty to protect that data. If behooves them to understand the implications of not encrypting data, or making the effort to understand the basics of such encryption in order to use it safely and successfully.

[1] And this is why digital-rights management by way of encryption will always fail. Ha, ha, ha.

In conclusion, we have developed a portable encryption solution for NetLogo that can be tailored to the various security requirements of either institutions or individuals. This has been done with the minimum coding required in order to reduce an possibility of mistake, and using pre-existing standard software that can utilise existing encryption infrastructure.

Acknowledgements

Bibliography