Security for R

This doc is a place for collaborative note taking and adding your questions during the call. This call will be recorded and posted within two days at https://ropensci.org/commcalls under “Past Calls”.

Background information on today’s topic, with speaker bios:
https://ropensci.org/blog/2019/04/09/commcall-may2019/

Agenda

1. Stefanie Butland - welcome (3 min)
2. Bob Rudis - Zen and the art of ensuring confidentiality & integrity in analytics workflows (20 min)
3. Ildi Czeller - Authorization vs authentication explained through signing commits: why you should do it and how ropsec helps you do it the right way (15 min)
4. Q & A (moderated by Bob Rudis, 20ish min)

Speakers’ slides and code

Bob’s slides available as Keynote, PDF and HTML from 3 sources:

- SourceHut
- GitLab
- GitHub

Ildi’s materials:

- slides

- ropsec package homepage
- vignette about commit signing
- Ildi will do a live demo. Folks interested in following along should install {ropsec} from GitHub with `devtools::install_github("ropenscilabs/ropsec")`
- Depending on your system you might have to install some system level dependencies, so it is worth looking at the vignette for this

Lots of R-centric and dev/datasci ecosystem security resources from Bob.

Participants (please add yourself here):

- *Name, Affiliation*
- Stefanie Butland, rOpenSci
- Bob Rudis, Rapid7
- Ildi Czeller, Emarsys
- Noam Ross, rOpenSci / EcoHealth Alliance
- Hugo Gruson, Université de Montpellier
- Brandon Hurr, Syngenta Vegetables
- Jenny Bryan, rOpenSci and RStudio
- David Severski, Starbucks
- Maëlle Salmon, rOpenSci
- Roel Hogervorst, Coolblue
- Jim Tyhurst, Tyhurst Technology Group
- Robyn Thiessen-Bock, NCEAS
- Mara Averick, RStudio
- John Smith, Shambhala
- Frie Preu, codecentric AG / CorrelAid
- Amanda Devine, Smithsonian Institution
- Andy Teucher, B.C. Ministry of Environment
- Kara Woo, Sage Bionetworks
- Apoorv Anand, CivicDataLab
- Scott Chamberlain, ropensci
- Stephen Froehlich, CableLabs
- Brad Sjue, Cerner
- Ju Kim, Fred Hutch
- Ellis Hughes, Fred Hutch
- Cody Marquart, Epistemic Analytics / UW Madison
- Helen Miller, Fred Hutch
- Julia Gustavsen, SOPHiA GENETICS
- Matt Mulvahill, Charter Comm.
- Nistara Randhawa, UC Davis
- Kate Webbink, Field Museum, Chicago
- Monica Gerber, Fred Hutch

- Anna Krystalli, University of Sheffield RSE
- Tim Golden, Morrison & Foerster LLP
- Julia Silge, Stack Overflow
- Sandor Felho, TransUnion LLC
- Sina Rüeger, EPFL
- Gavin Simpson, University of Regina
- Jeroen Ooms, rOpenSci
- Rich FitzJohn, rOpenSci / Imperial College London
- + 5 more participants not signed in

Questions for speakers (add yours here as you think of them):

Please include your name with your question. The Moderator will call on you to ask your question, unless you note you prefer to have the moderator ask on your behalf

- (name) question
- (David S) Are there any (good) options to verify the integrity of code from either CRAN or GitH__? Given that CRAN sometimes alters source code as part of the submission/publication process, this has been a sticky issue to wrangle.
    - Bob: The best option is a local CRAN repo. If you can afford it, RStudio offers this as part of RStudio-Connect.
- (David S) Can you talk about the threat vectors that code signing in SCM can/may/cannot address? I sign my commits but am not sure what it gains me (or the consumers of my code).
    - Maybe a dtupdate enhancement to scan a source repo and see if the commits are signed and valid? :)
    - I can possibly add something to https://github.com/ropenscilabs/defender (will try to file an issue)
- (Frie Preu) Is there a way to use R how to encrypt whole folders? Kind of as an alternative to tools like Veracrypt? We do pro bono data science consulting for NPOs and we need a way to share and locally store the (potentially sensitive) data within our project teams.
    - Can use {cyphr} to decrypt/re-encrypt around analyses. Can use {keybase} or just Keybase alone to do this as well.
- (Frie Preu) Any comments on sharing keepass files (kpbx) via the cloud with a team? We can't really afford a cloud-hosted password manager. (hrbrmstr) https://keepass.info/help/base/multiuser.html talks abt this a bit. I'd suggest gdrive/dropbox/onedrive/keybase as a method for sync'ing them -> thanks! :)
    - LastPass (https://www.lastpass.com/) is the one that's free and I know supports separate shared (group/team) vaults

- (David S) Any (non-binding) recommendations (or anti-recommendations) for survey sites that pass the bar for handling responses safely?
    - (hrbrmstr) When we had to source data from breached third parties back in my Verizon DBIR days we used https://www.surveygizmo.com/ but more important than the platform is what fields you download from the actual survey itself. I'll explain more on the Q&A & type in a bit. So I didn't explain more but the gist is only store the fields you need. Leave out IP address, gender, etc. Def bob@rud.is me if you wld like an extended discussion.
- (Noam R) What is a recommended approach to running CI when a workflow requires sensitive/confidential information that is larger than can fit in environment variables? Currently using rcrypt to save this info and decrypting with a key in an env var. Seems shaky. 😬
    - Bob: That's the hack, but it is a limited hack. For confidential info you should run on a local containerized CI service so you don't have to send stuff to servers you don't control.
    - Bob: https://drone.io/ is what you can use with something like gitea (which is what I'd recommend for local git). You can pretty easily (i don't use that word lightly) setup a local git config to push and pull to/from multiple sources (I do it all the time) and then use drone locally (it's all docker based) for any seekrits that won't easily work externally or in public CIs. At some point that's a necessity (tho, again, your hack is excellent).
- (Mara A) Any good recommendations for troubleshooting "Invalid crypto engine" errors? 😳
    - Ildi: 1 thing I suspect is gpg <2 and >2 versions, you should probably have >2, but I am not sure
    - Mara: Thx! I think I have gpg 2.2.10, so I'll do a bit more digging! (Basically I'm scared to break my current gpg setup by changing things up to get ropsec running)
- (Nistara R) What minimum OS security options should we make sure to fix. You mentioned a strong OS password
    - Bob mentioned turning on disk encryption at the top of his talk. (Noam). Thanks! (Nistara)
    - Also check out: https://github.com/ropenscilabs/ropsec#lightweight-system-checks ({ropsec} has a function which will provide platform-specific guidance
- (Ellis H) Does the 'verified' functionality that you get with ropsec only work with github? Or is it valid in any git workflow?
    - It works with any git flow. ropsec supports gitlab out of the box as well, with other repository management services you have to upload your public key manually. Even without a centralized service if you share your public key, anyone has the ability to verify those commits. (though ropsec does not have wrappers for this). Git did not always supported gpg, but is supports it for a while now. You can test verification locally from the command line with git verify-commit command

Add your notes and resources here:

Have a favorite screencast, tutorial, blog post about these tools or their applications? Note them here so we can share with others.

We (Rich FitzJohn) have a new package for working with Hashicorp's vault for secrets-as-a-service: https://github.com/vimc/vaultr - hopefully will end up on CRAN later this year. We've been using this in production for a couple of years now (sorry for the late breaking spam)

What would you like to hear about on a future rOpenSci Community Call?

Upcoming Call - How do organizations include and involve multilingual communities? - with Emilio Bruna, Rayna Harris, moderated by Melina Vidoni - early July 2019 Details will be posted at https://ropensci.org/commcalls/

Ideas we're considering

- Reproducible research / modern reproducible data science
- Geospatial analysis in R with Antarctic & Southern Ocean research as example
- Make the most of GitHub for package development and R ecosystem knowledge
- Live code review for advanced developers
- Maintaining a package
- rOpenSci tools to access academic literature
- Text processing & analysis featuring rOpenSci tools

Tell us here what would you like to hear about on a future rOpenSci Community Call