ON THE NEW (AND NOT-SO NEW) CLAIMS ABOUT GUCCIFER 2.0

The initial files released by the persona Guccifer 2.0 on June 15, 2016 included — in addition to graffiti paying tribute to Felix Dzerzhinsky, the founder of Russia's secret police — metadata deliberately set to Cyrillic (the metadata had previously been interpreted, implausibly even at the time, to be a mistake).

And a file later released on September 13, 2016 purportedly from Guccifer 2.0 but released via a magnet site and never linked on his WordPress site, was probably copied, locally, to a Linux drive somewhere in the Eastern time zone on July 5, 2016; the files were then copied to a Windows file on September 1, 2016.

Those are the fairly uncontroversial findings from two separate research efforts that have recently renewed debate over whether the conclusion of the intelligence community, that Russia hacked the DNC, is valid.

I'm going to do a two part post on this issue.

What to Read

As you might be able to figure out, nothing about those two conclusions at all dictates that the Intelligence Community conclusions that Russia is behind the hack of Democratic targets are wrong. The reason they're so controversial is because they've been used, in tandem, to support claims that the IC conclusion is wrong, first in a (to me) unconvincing letter by the Veteran Intelligence Professionals for Sanity (chiefly Bill Binney, Kirk Wiebe, Ed Loomis, and Ray McGovern), and then in some even sloppier versions, most notably at the Nation. In between the original analysis and these reports are some other pieces making conclusions about the research itself that are in no way dictated by

the research.

In other words, it's all a big game of telephone, some research going in the front end and a significantly distorted message coming out the back end.

So before I get into what the two studies do show, let's talk about what you should read. The first argument has been made by Adam Carter at his G2-space, which is laudable as a resource for documents on Guccifer 2.0, no matter what you think of his conclusions. There's a ton in there, not all of which I find as persuasive as the argument pertaining to the Russian metadata. Happily, he made two free-standing posts demonstrating the RSID analysis (one, two). I first discussed this analysis here.

The RSID analysis showing that the cyrillic in Guccifer 2.0's documents was actually intentional relies, in part, on the work of someone else, posting under the name /u/tvor_22. His post on this is worthwhile not just for the way it maps out how people came to be fooled by the analysis, but for the five alternative explanations he offers. In in no way think those five possibilities are comprehensive, but I appreciate the effort to remain open about what conclusions might be drawn from the evidence.

Between those three posts, they show that the first five documents released by Guccifer 2.0 were all copied into one with certain settings set, deliberately, to the Russian language. That's the first conclusion.

The forensics on copying was done by a guy posting under the name The Forensicator, whose main post is here. Note his site engages in good faith with the rebuttals he has gotten, so poke around and see how he responds. He argues a bunch of things, most notably that the first copy of files released in September was copied locally back in July, perhaps from a computer networked to the host server. That analysis doesn't rule out that the data was on some server outside of the DNC. I raised one concern

about this analysis here.

Finally, for a more measured skeptical take — from someone also associated with VIPS who did not join in their letter — see Scott Ritter's take. I don't agree with all of that either, but I think a second skeptical view is worthwhile.

All of which is to say if you want to read the analysis — rather than conclusions that I think go well beyond the analysis — read the analysis. Assuming both are valid (again, I think the RSID case is stronger than the copying one), the sole conclusions I'd draw from them is that the Guccifer 2.0 figure wanted to be perceived as a Russian — something he succeeded in doing through far more than just metadata, though the predispositions of researchers and the press certainly made it easy for him. And, some entity that *may* associated with Guccifer 2.0 (but may also be a proxy) is probably in the Eastern Time Zone, possibly (though not definitely) close to the DNC (or some other target server). That's it. That's what you need to explain if you believe both pieces of analysis.

Whatever explanation you use to explain the inclusion of Iron Felix in the documents (which is consistent with graffiti left *in* the hacked servers) would be the same one you use to explain why the metadata was set to Cyrillic; the IC and people close to the hack have explained that the hackers liked to boast. And the only explanation you need for the local copy is that someone associated with the Russians was close to DC, such as at the Maryland compound that got shut down.

Guccifer and the DNC … or DCCC … or Hillary

Since we're examining these claims, there's another part of the presentation on the RSID data (and Carter's site generally), that deserves far more prominent mention than the current debate has given, because it undermines

the framing of the debate. We've been arguing for a year about Russia's tie to Guccifer 2.0 based on the persona's claim to have provided DNC documents to WikiLeaks. **But the documents originally released in the initial weeks by Guccifer 2.0 were, by and large, *not* DNC documents**. As far as I know/u/tvor_22 was the first to note this. He describes that the Trump document first leaked only appears via other sources as an attachment to a Podesta email, though there are alterations in the metadata, as are three of the others, with the fifth coming from an unidentified source.

> Let's take the very first document posted by Guccifer2.0, which some security researchers have cited as 'an altered document not properly sanitised.' If we diff the raw copy – pasted into text documents – of both the original Trump document found in the Podesta emails and the Guccifer 2.0 version, ignoring white-spaces and tabs (diff -w original.txt altered.txt):
>
> - *the table of contents has been re-factored.*
> - *many of the links are naked in the Guccifer2.0 version. (Naked as in not properly behind link titles, indicating Guccifer2.0's version may have been an earlier draft.)*
> - *the error messages are in Russian.*
> - *None of the above quirks could be found in comparing 2,3, or 5.doc to their*

> *originals (100% textually equivalent). 4.doc could not be found on WikiLeaks for a comparison.*
>
> **None of the textual content in any of these four 'poorly sanitised' documents has been altered, removed, or doctored.** In other words all the differences you would expect from a copy and paste from one editor to another. So why bother copy and pasting into a new document at all? I wonder.
>
> [1.*doc's* ***original***, *2.doc's* ***original***, *3.doc's* ***original***, 5.doc's ***original***. *4.doc could not be found in Wikileaks. The bare texts of 2,3, and 5 are checksum equivalent.*]

G2-space has posted an expansion of this analysis, by JimmysLlama. It provides a list for where the first 40 documents (covering Guccifer 2.0's first two WordPress posts) can — or cannot — be found. The source for (roughly) half remains unidentified, the other half came from Podesta's emails. At the very least, that reporting makes it clear that even for documents claimed (falsely) to be DNC documents, Guccifer had a broader range of documents than what WikiLeaks published.

That explains reporting from last summer that indicated the FBI wasn't sure if WikiLeaks' documents had come from Russia/Guccifer 2.0.

> The bureau is trying to determine whether the emails obtained by the Russians are the same ones that appeared on the website of the anti-secrecy group WikiLeaks on Friday, setting off a firestorm that roiled the party in the lead-up to the convention.
>
> The FBI is also examining whether APT 28

> or an affiliated group passed those emails to WikiLeaks, law enforcement sources said.

Now we know why: because they weren't the same set of files as had been taken from the DNC (though the FBI did already know some Hillary staffers had been hacked.) See this post from last summer, in which I explore that and related questions.

The detail that Guccifer 2.0 was actual posting Hillary, not DNC, documents is somewhat consistent with what John Podesta has said. He revealed that he recognized an early "DNC" document probably came from his email.

> And other campaign officials also had their emails divulge earlier than October 7th. But in one of those D.N.C. dumps, there was a document that appeared to me was— that appeared came— might have come from my account.

Podesta he has always been squirrelly about thus stuff and probably has reason to hide that the Democrats' claims that Guccifer 2.0 was releasing DNC documents were wrong (indeed, *that's* something that would be far more supportive of skeptics' alternative theories than this Guccifer 2.0 data, but it's also easily explained by Democrats' understandable choices to minimize their exposure last summer). Importantly, Podesta also suggests that "other campaign officials also had their emails divulged earlier than October 7th," without any suggestion that that is just via DC Leaks.

On top of a lot of other implications of this, it shifts the entire debate about whether Guccifer 2.0 was WikiLeaks' source, which has *always* focused on whether the documents leaked on July 22 came from Guccifer 2.0. Regardless of what you might conclude about that, it shifts the question to whether the Podesta emails WikiLeaks posted came from

Guccifer 2.0, because those are the ones where there's clear overlap. Russia's role in hacking Podesta has always been easier to show than its role in hacking the DNC.

It also shifts the focus away from whether FBI obtained enough details from the DNC server via the forensic image it received from Crowdstrike to adequately assess the culprit. Both the DNC and Hillary (as well as the DCCC) servers are important. Though those that squawk about this always seem to miss that FBI, via FireEye, disagreed with Crowdstrike on a key point: the degree to which the two separate sets of hackers coordinated in targeted servers; I've been told by someone with independent knowledge that the FBI read is the correct one, so FBI certainly did their own assessment of the forensics and may have obtained more accurate results than Crowdstrike (I've noted elsewhere that public IC statements make it clear that not all public reports on the Russian hacks are correct).

In other words, given that the files that Guccifer 2.0 first leaked actually preempted WikiLeaks' release of those files by four months, what you'd need to show about the DNC file leaks is something entirely different than what has been shown.

New Yorker's analysis on coordination

That's a task Raffi Khatchadourian took on, using an analysis of what got published when, to argue that Russia is WikiLeaks' source in his recent profile of Assange (I don't agree with all his logical steps, particularly his treatment of the relationship between Guccifer 2.0 and DC Leaks, but in general my disagreements don't affect his analysis about Russia).

> Throughout June, as WikiLeaks staff worked on the e-mails, the persona had made frequent efforts to keep the D.N.C.

> leaks in the news, but also appeared to leave space for Assange by refraining from publishing anything that he had. On June 17th, the editor of the Smoking Gun asked Guccifer 2.0 if Assange would publish the same material it was then doling out. “I gave WikiLeaks the greater part of the files, but saved some for myself,” it replied. “Don’t worry everything you receive is exclusive.” The claim at that time was true. None of the first forty documents posted on WordPress can be found in the WikiLeaks trove; in fact, at least half of them do not even appear to be from the D.N.C., despite the way they were advertised.
>
> But then, on July 6th, just before Guccifer 2.0 complained that WikiLeaks was “playing for time,” this pattern of behavior abruptly reversed itself. “I have a new bunch of docs from the DNC server for you,” the persona wrote on WordPress. The files were utterly lacking in news value, and had no connection to one another—except that *every item* was an attachment in the D.N.C. e-mails that WikiLeaks had. The shift had the appearance of a threat. If Russian intelligence officers were inclined to indicate impatience, this was a way to do it.
>
> On July 18th, the day Assange originally planned to publish, Guccifer 2.0 released another batch of so-called D.N.C. documents, this time to Joe Uchill, of *The Hill*. Four days later, after WikiLeaks began to release its D.N.C. archive, Uchill reached out to Guccifer 2.0 for comment. The reply was “At last!”
>
> [snip]
>
> Whatever one thinks of Assange’s election disclosures, accepting his

> contention that they shared no ties with the two Russian fronts requires willful blindness. Guccifer 2.0's handlers predicted the WikiLeaks D.N.C. release. They demonstrated inside knowledge that Assange was struggling to get it out on time. And they proved, incontrovertibly, that they had privileged access to D.N.C. documents that appeared nowhere else publicly, other than in WikiLeaks publications. The twenty thousand or so D.N.C. e-mails that WikiLeaks published were extracted from ten compromised e-mail accounts, and all but one of the people who used those accounts worked in just two departments: finance and strategic communications. (The single exception belonged to a researcher who worked extensively with communications.) All the D.N.C. documents that Guccifer 2.0 released appeared to come from those same two departments.
>
> The Podesta e-mails only make the connections between WikiLeaks and Russia appear stronger. Nearly half of the first forty documents that Guccifer 2.0 published can be found as attachments among the Podesta e-mails that WikiLeaks later published. Moreover, all of the hacked election e-mails on DCLeaks appeared to come from Clinton staffers who used Gmail, and of course Podesta was a Clinton staffer who used Gmail. The phishing attacks that targeted all of the staffers in the spring, and that targeted Podesta, are forensically linked; they originated from a single identifiable cybermechanism, like form letters from the same typewriter. SecureWorks, a cybersecurity firm with no ties to the Democratic Party, made this assessment, and it is uncontested.

Now, I'd like to see the analysis behind this publicly. It should be expanded to include all

the documents leaked by Guccifer 2.0. It should include more careful analysis of the forensics behind the phishes (security companies have done this, but have not shown all their work). Moreover, it doesn't rule out a piggyback hack, though given that Guccifer 2.0 was leaking Hillary emails from the start, it's unclear how that piggyback would work. All that said, it provides a circumstantial case that these were the same two sets of documents.

Khatchadourian doesn't dwell on something he alluded to here, which is that all the DNC documents were email focused, collected from just 10 mailboxes. That's the nugget that, I suspect, Assange will point to (and may have shared with Dana Rohrabacher) in an effort to rebut the claims his source was Russia (one thing Khatchadourian gets wrong is what Craig Murray said about two different sources for WikiLeaks, but then he points to a WikiLeaks claim they got the emails in late summer and September 19 date on all of them — not long before Murray picked something up in DC — so that's another area worth greater focus). For now, I'll bracket that, but while I suspect it points to really interesting conclusions, I don't think it necessarily undermines the claim that Russia was Assange's source. More importantly, none of the things people are pointing to in this new analysis — the metadata in files released by Guccifer 2.0, the metadata in files released on a magnet site but never directly by Guccifer 2.0 — affects the analysis of how completely unrelated emails got to WikiLeaks at all.

All of which is to say that the these two pieces of analysis actually miss the far more interesting analysis that got done with it.

Update: Turns out the Nation issued a correction today, which reads in part,

> Subsequently, *Nation* editors themselves raised questions about the editorial process that preceded the publication of the article. The article was indeed

> fact-checked to ensure that Patrick Lawrence, a regular *Nation* contributor, accurately reported the VIPS analysis and conclusions, which he did. As part of the editing process, however, we should have made certain that several of the article's conclusions were presented as possibilities, not as certainties. And given the technical complexity of the material, we would have benefited from bringing on an independent expert to conduct a rigorous review of the VIPS technical claims.

It added an outside analysis by Nathanial Freitas of the two reports, a rebuttal from VIPS members who did not join the letter, and a response from those who did. Freitas provides a number of other possibilities to get the throughput observed by Forensicator. The VIPS dissenters raise some of the same points I do, including that this server may be somewhere outside of DNC.

> It is important to note that it's equally plausible that the cited July 5, 2016, event was carried out on a server separate from the DNC or elsewhere, and with data previously copied, transferred, or even exfiltrated from the DNC.
>
> However, independent of transfer/copy speeds, if the data was not on the DNC server on July 5, 2016, then none of this VIPS analysis matters (including the categorically stated fact that the local copy was acquired by an insider) and simply undermines the credibility of any and all analysis in the VIPS memo when joined with this flawed predicate.