039| Deconstructing the Dukes: A Researcher's Retrospective of APT29

blog.f-secure.com/podcast-dukes-apt29/

May 6, 2020

APT29, aka Cozy Bear or the Dukes, is a cyber espionage group whose misdeeds include hacking the DNC servers in the runup to the 2016 US election. Now, as the subject of MITRE's latest ATT&CK Evaluation, the group is in focus again. The Dukes are familiar to F-Secure's Artturi Lehtio, who extensively researched them during their heyday and authored a 2015 whitepaper on the topic. But hindsight is 20/20 and Artturi joins us for Episode 39 of Cyber Security Sauna to discuss how, over time, his views on the group have evolved.

Also in this episode: What APT groups usually do after being burned and how the Dukes is different; why calling them a "single organization" is too strong; and why published research into APT groups has generally dwindled in recent years.

Listen, or read on for the transcript. And don't forget to subscribe, rate and review!

Janne: Welcome, Artturi.

Artturi: Hey Janne, thanks for having me.

Absolutely. We're talking about Dukes because APT29 is the scenario MITRE ATT&CK chose for round 2 testing this year, and you're the guy who led F-Secure's investigation into Dukes in 2015. What are you working on these days?

These days I'm working on totally unrelated topics on business strategy and corporate development at F-Secure. But yeah, back in the day I spent a few years researching these kinds of APTs.

So you're all grown up.

I'd like to think so.

Okay. Can you give us a quick rundown on the Dukes, or APT29. Who are they, and who do they work for?

So the Dukes, APT29, Cozy Bear, are all different names for the same set of tools and activities used for cyber espionage that we believe are sponsored by the Russian Federation. It's a set of tools and activities that have been in operation since at least 2008 up to, at the time of the report, up to 2015.

Artturi Lehtio

All right. You're calling them Dukes. What's wrong with Cozy Bear?

The first public mention of the Dukes was in February 2013 when Kaspersky, together with an organization called CrySyS Lab, which is part of the Budapest University of Technology and Economics. They released a whitepaper on a specific piece of malware that they named MiniDuke. And later then when we started discovering other toolsets that were clearly

related, us and some others in the industry started giving them names that would always end in Duke, to show that there was some relationship between them. So then CosmicDuke came later, and CozyDuke eventually as well.

Cozy Bear was a name that was used by a specific company in the industry later, and Cozy Bear and Cozy Duke, in terms of toolsets, are essentially the same thing, but Cozy Bear can also then refer to the wider set of activities. We talked about Cozy Bear internally for a while and actually changed the name to CozyDuke when we realized it's related to the Dukes, just to keep it in our opinion much simpler and make it clear that there's a relationship. But some have stuck to using Cozy Bear.

Okay. So like you said, the group was active up to the time when you published your research. What happened with them since?

So yeah, that's a really interesting question actually. Most of the time when new research is published on tools or techniques used by threat actors, the actors will try and make the minimum amount of changes needed just to be able to continue their operations. So they'll make some small modifications to tools, or they'll set up some new infrastructure or something, but try and keep going as soon as possible. And the Dukes did that for a long time as well, for many years. And within weeks of new research being published, they'd be back at it with the slightest of modifications.

That's why it was really interesting actually, after we released the whitepaper in September of 2015, they really did seem to go dark for about half a year, at least. And only later in the spring of 2016 there started to be small signs of them starting to set up new infrastructure again. And then there were smaller outbreaks of activity in 2016 and '17 and '18 where they did some activities using slightly different tool sets, but they never really came back in the same force that they used to operate in. Or at least, not in the same way that we would still classify them as the same set of activities.

That's interesting. What do you think happened to the people and tools and operations?

The people and the know-how lived on, I'm sure of that. Whether the people were moved to other organizations or just switched tools and techniques or so forth, I don't know. They probably retired a lot of the tools because at that point the tools were so well known, and at that point they'd already started to get into the habit of taking into use new tool sets at an accelerating pace. So they were increasingly diversifying already in terms of tool sets. So they probably continued that as well.

So is that typically what happens when a group like this gets burned? Do they go away and take a breather, and rethink about their operations, retool, and come back, version 2.0? Or how does that look?

Actually, not usually. Most of the time they'll try and continue, and that's probably for a couple of reasons.

One reason, definitely, is for these types of cyber espionage capabilities, very few nations have the resources where they can just shut down one of their main capabilities without losing a significant amount of intelligence. So they have to make that cost benefit analysis of, is it worth losing that capability? It takes time to rebuild. And so you have to be happy with not having that capability for a while.

Another reason is also the amount of effort involved, and then it depends on what your objectives are. In the case of the Dukes, they were always about information gathering. So for them, being able to maintain a presence in an organization's networks, preferably for years, just collecting information, was important. And once you're burned bad, it's just not something you can continue doing anymore. For other actors, for instance Fancy Bear or APT28, another suspected Russian espionage group, they do a lot more active measures. They do disruptive attacks. They collect information for information warfare purposes or for doing information leaks, for instance. Those are much more short-lived and so it's a different objective, and therefore it's a different cost benefit analysis as well.

So is that the reason of doing research like this and publishing it? To burn these groups? Or what's your motivation and purpose for doing this kind of research?

That's not the primary purpose. The ultimate purpose of really any cyber security research, and really any APT research, is to enable defenders to make more informed decisions. So, to help defenders prioritize their activities, prioritize what types of defensive capabilities they build, and how they build those capabilities.

So when we do this type of research, it's for that purpose as well. It can be to dive really deep into a technical analysis of a specific tool or set of tools or a specific technique so that defenders have a better chance of preparing for that and defending against it. Or it can be something like we did with the 2015 whitepaper, where it was more a high level overview and a look at how the tactics had evolved over years, and we tried to look at behavioral patterns and how those patterns had changed over the years to help defenders then assess how the threat landscape may be evolving, how attackers may be evolving, and how to prioritize future investments and defensive measures.

Now that you have the benefit of hindsight, what do you think you might have done differently while conducting or publishing your research? Like, if you could do it all over again?

I think the biggest challenge for this kind of research is knowing where to stop or being willing to stop at some point. As a researcher you keep finding new threads and keep coming up with additional hypotheses and you keep wanting to go just a bit further before you stop and write the report. But at some point you just have to stop, and you have to kind of accept

that it's a snapshot. It's your best understanding at that point in time, but it can't be more than that. And so I think we did a relatively good job in terms of explicitly bringing out the data we had and the analysis we did based on that data, and also the limitations of it.

At the same time, I think there are a couple of assessments where my thinking has developed further. I think the most important evolution is in the report, we state, actually quite strongly, that we believe the Dukes to be a single large well-coordinated organization. In hindsight, I would soften that statement a bit. I still believe in the underlying sentiment. I still believe in the data and how it shows a surprisingly significant amount of coordination across all of the different activities. But I think the "single organization" is maybe a bit too strong, in that there's probably multiple operational groups, multiple separate teams of operators using the tools. There's probably multiple separate teams of developers developing different parts of the tool sets. And then those tool sets get distributed to the different operational teams at slightly different points in time, or some teams get some tools and others get a different set of tools, and the groups may coordinate on some things or share some knowledge, but not everything.

So you're almost describing a company, an organization, with the developers and task masters, but then separate teams working on separate objectives, people moving between the teams sharing some of the tools, sharing some of the targeting. Is that what you think is happening?

Yeah. I'll give a couple of examples. So, there's probably multiple operational teams. Because when we looked at the specific ways some of the tools were used or configured, for instance in the case of CosmicDuke, the same tool set, we could kind of see different preferences for what encryption keys to use, or how to configure it or how to use it. And

those would usually cluster around. So it seemed like different operators or different teams had different preferences, but then there was sometimes some overlap in some shared infrastructure or some similarity. So maybe a person moved from one team to another or something like that. Or in the case of some of the tools, we could see how a new version of the tool was created, and then these different clusters of different ways of using the tools. They'd take the new version into use at slightly different points in time, depending on maybe their operational cycle, for instance. They'd be at different stages of the operational cycle, so they'd wait until the start of the next cycle to update their tools, for instance.

But then in some situations we also saw cases where one tool set was used to deploy another tool set on the same system, in situations where the tool sets served the exact same purpose and have the same functionality, they're just different tool sets and connecting to different clusters of infrastructure. And the most likely explanation for that that we could come up with was probably handing off a target from one operational team to another one. Where the easiest way, perhaps, was to deploy the other team's implant onto that system and then remove your implant and they'd retain control of it.

So there were clearly situations where there was some collaboration happening or some sharing happening, but most of the time, they were operating relatively independently.

The reason we originally said it was a single coordinated organization was because we never really saw overlap in the sense where it would have looked like separate operational teams were targeting the same target at the same time, without knowing about each other. So clearly they did a lot of deconflicting.

In the targets as well, the majority of the activity was focused on foreign policy targets. But then there was a small set of activity that very much looked more like police law enforcement work, where the targets seemed to be criminals engaged in illicit drugs, growth hormones, and the sale of those, and actually, in child porn. So that seems very much like a law enforcement target. And those seemed to be targeted at individuals in Russia. And then you'd see the non-Russian foreign ministries get targeted by a separate cluster of activity, again suggesting that maybe there were different operational teams, maybe even in different organizations who were sharing some of the methodologies, some of the tools, but had their different priorities and worked relatively independently.

But does that also support your work towards attribution? I can think of a Russian organization whose mandate just happens to have in it both foreign espionage of neighboring countries, but also drug trade within Russia.

Well, I'm not an expert in criminology or Russian intelligence organizations. But certainly, for instance, the Estonian Foreign Intelligence Service, in their 2018 public report, they actually explicitly attribute the Dukes to Russia's FSB and SVR, the federal security service and foreign intelligence service respectively, suggesting that it was, or part of it may have been actually a joint operation by those two.

Is there anything else, sort of a hunch you got about the people behind these attacks when you were doing your research? Did you get the sense that they're sort of young, fresh, quick minds, or older, more seasoned veterans?

That's another interesting question. Not in the sense that we would have some idea of who the people were. But when we looked at the evolution of the different tool sets, like MiniDuke for instance, they look like they've been developed by people who have a lot of experience developing malware and viruses. MiniDuke famously, parts of it were written in assembly language. So very low level, very challenging to code in that sense, but something that, you know, the good old days virus writers would do.

Some other tool sets like CozyDuke, for instance, versions of it had a lot of helper functionality and debugger functionality and logging functionality, and it looked like you'd taken a team of enterprise software developers and told them to go and write malware, and they started from that background, and then slowly evolved into making it look more like malware, and they started removing some of the logging and started obfuscating things, and started hiding things eventually over the years.

So on the tool set side, it looks like different tool sets were developed by developers or teams of developers from very different backgrounds and those eventually kind of fed into each other.

That's interesting. Getting back to your research, what kind of lessons can be learned about where you were successful and where less so?

I think it continues to show the importance of really making the data and the analysis explicit, so others can then further build on that work, and others can make their own assessments and decide to agree or decide to disagree. I think it's an important but difficult balance to strike between analysis and over-interpretation, of analyzing and presenting hypotheses or even conclusions, but not over-interpreting, reading too much into it.

I think it's incredibly important to build on the research that others have done. We made the conscious decision of citing others as much as possible and putting in a long list of references and links to further reading, more technical analyses of the different tool sets that others have done, rather than just try to gain a short term marketing or PR win by just emphasizing novelty or trying to make it look like it was just our research and just our findings.

Yeah, you mentioned other companies who've been doing research and publishing research around the same group. Do you find that that makes it harder to look at things with fresh eyes and keep an open mind when making conclusions, or is it more helpful because they sort of fill in some of your blanks?

It's absolutely a great benefit and a positive thing. Any researcher, any organization has a limited point of view, limited visibility in that sense, and they only see a portion of the activity. I never had the opportunity to do actual live incident response for instance, on a Dukes incident. We based our research primarily on observing tools and infrastructure and artifacts we could find, and then inferring from that how those tools might have been used, whereas, you know, others have then later gone on to talk about their perspective from doing incident response.

Like in 2016, there was a talk by Matthew Dunwoody and Nick Carr, who were working for Mandiant at the time, a talk called No Easy Breach, and they talked about their experiences responding to a single Dukes incident. And that was really interesting because that was very much the opposite perspective.

Costin Raiu, who leads the team at Kaspersky that does APT research, he's famously compared APT research to paleontology, where you study animals and plants from bones and fossils. And kind of looking at a pile of bones and trying to figure out what the animal looked like and how it moved and what it did and what it ate and so forth. And our research in the Dukes was very similar to that in that we had all of these small artifacts, kind of bones in a warehouse, and we tried to figure out what the animals might have looked like, where the different bones belonged, but we hadn't seen any of those animals actually live or moving.

And then Dunwoody and Carr, they'd done the incident response, they'd seen some of those animals live in their natural habitat. They'd seen some of the activities, and so they talked about that perspective. And it was really interesting to compare how far we'd gotten in our assessment from our perspective compared to their perspective and how well or not we'd been able to make those inferences, for instance, about how the tools were being used.

Do you ever get that feeling where you're reading some additional research and you're looking at a conclusion and you're like, "Oh man, that's why they did that!" or "That's what my data showed back then!" but you never made that connection, and you're like, "I should have caught that."

There's a couple of examples, actually. One is HammerDuke or Hammertoss, which FireEye released a whitepaper, I think it was in August 2015 on what they called Hammertoss, when they also first used the name APT29, and they talked about this super stealthy, super advanced piece of malware that they'd found from this novel threat actor called APT29.

In fact, at that point, we'd seen that tool a lot, and we'd looked into it, but from our perspective, we'd always just seen it as being used later in an attack as kind of a helper or utility. They'd be using other malware earlier in the breach and then they'd maybe drop HammerDuke on some of those systems and leave it there as kind of a secondary tool. So we didn't pay much attention to it.

And then, from FireEye's perspective, I assume they found it in an incident response investigation, where some of the earlier tools had already been removed from the system. So for them, it was the star of the show. So that was a very different perspective.

Another, more recent, was late last autumn, ESET released research on what they called Operation Ghost, on activity that again, they tied to the Dukes of old. But actually, based on a quick look, it's related to some of the tool sets and some of the pieces of activity that we never had as much time to dive into, some of the OnionDuke stuff where we were left with some open questions. And I never had a chance to go as deep into that as I maybe would have wanted. So it was really interesting to see someone else had then managed to find out a lot more about that side of it, that again, we never actually learned as much about.

So yeah, that happens quite a lot, where the research that others then release brings new perspectives or new ideas or fills in the blanks, like you said, where you realize, "Now that makes sense," or "That's why they did it," or "That's how it's related," and so forth.

Okay. I mean, you're referencing some fairly recent research there. But it seems to me like we're not seeing a whole lot of APT research anymore. Do you think that's accurate?

I do think it's accurate. It's much rarer now. I think there was a point in time when APT research and making it publicly available was a hot trend. Unfortunately, an increasing number of people and companies got into it primarily for short term marketing and PR winds, you know, trying to do some research and write a quick report and get quoted by the media and that way get your company name out there. And that kind of led to a race to the bottom, where you needed to constantly come up with ever more sensational headlines and conclusions.

And the problem was, most of the time, most of the research was looking at a small set of actors and looking at the same sets of data, the same data sources, so most of the people were actually looking at the same things but trying to turn it into new sensational headlines and novel research. And that led to eventually kind of a bubble where everyone just started emphasizing novelty and exaggerating the novelty and stopped referencing others and stopped giving context. And smaller and smaller chunks of activity or tools became the next greatest piece of new findings.

And eventually, the research lost its value for defenders. Because from a defender's perspective, if you lose the context of what's this related to and what else should I be looking into, and what else should I care about, and if everything's the most advanced, and everything's the most stealthy and the most novel, then how do you prioritize as a defender anymore? Everything is the most at everything, and suddenly everything's equal again. So as a defender, you can't really use the information to prioritize anymore.

So it lost value, and – I think there was a bubble – I think the bubble burst. Some of the researchers and companies who were good at doing original research, they started releasing less of it publicly and turned it into a business and started charging money for it. And others figured out new ways of doing marketing and PR for them. It had many benefits, but one of the downsides was less public research.

I get that as a researcher, you're not a big fan of the sensationalism of it all. But do you think we're worse off when we don't have that equally advanced, equally persistent effort from researchers towards the APT groups?

I wouldn't draw that conclusion. I think there's still more research being put into these types of topics.

It's just not being published.

Exactly. I think the research kind of returned to its roots. From becoming for a while something that was driven by marketing and PR, it went back to – you know, these days, incident responders are doing a lot of research and using a lot of research. Red teamers are doing a lot of this kind of research. Cyber security product R&D organizations are doing and using this type of research, in-house threat intelligence teams exist at more well-resourced defenders and so forth. So the research is being moved back to where it brings more value, back where, in my opinion, it belongs, informing defenders and improving defenses, rather than being primarily a marketing activity.

Okay. Now, if we think about specifically the hacks into the DNC servers that preceded the 2016 election, what do you think was the purpose of the Russian government in doing that?

So you're talking about the reports in June 2016 that there was an ongoing incident response and investigation into IT systems of the US Democratic National Committee, the DNC.

Yeah. "Show me the server!"

It was before the "show me the server," but yeah, same case.

Oh, was it? Okay.

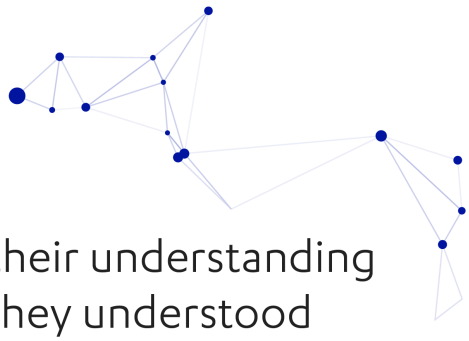
So in June 2016, CrowdStrike, the US company doing the incident response, went public on the fact that they were doing that incident response and some of their early findings. The blog post is fun to read, kind of because of that "find me the server" instance or case, because they've had to update it many times over the years. I think the most recent update is actually from January this year, where they added again links to new stories refuting the "find me the server" conspiracy theory. So it's caused them a lot of trouble as well.

So, June 2016, CrowdStrike went public and said they'd identified two separate breaches by two separate threat actors in the IT systems of the Democratic National Committee. One of the actors was the Dukes, or, as CrowdStrike called them, Cozy Bear. The other actor was Sofacy, also know as APT28 or Fancy Bear.

And what CrowdStrike also stated then was that they hadn't actually identified any collaboration between the two actors, and in fact, they hadn't identified any signs of even awareness by the two actors of one another, which is actually quite significant information.

What they do also say is that they suspect the breach by the Dukes began in the summer of 2015 already, whereas APT28 only entered the systems in April 2016. Which again, is actually important from a timing perspective because in the summer of 2015, that was still before the whitepaper for instance, and back when the Dukes were still actively targeting new organizations, especially targets relevant for foreign policy goals.

And in the case of the DNC I think they were there just to gather information. I think they were there just because it's an important political target if you want to understand how the foreign policy and security policy of the US may evolve in the future. So they wanted to get that visibility, they wanted to be there just listening in, gathering information.



> "A lot of the Dukes' finesse was their understanding of humans and organizations. They understood how incident response proceeds and how to make it harder, how defenders work and what their weaknesses are."
>
> **-ARTTURI LEHTIÖ, F-SECURE**

Well as an information-gathering operation I'm pretty sure that that counts as a success. So do you think everything else sort of just was opportunistic, building on that initial success?

I guess you're referring to some of the data leaks, for instance…

Yeah. Information operations, all of that.

So again, that's why it's important to be aware that there were two separate breaches and two separate actors that operated independent of each other. The information leaks and all of the hacking the election that's been talked about, that's all been tied to APT28, the other actor. Not APT29, not the Dukes.

So to the best of my understanding, the Dukes were there to gather information, and then the other kids on the block walked in and started making a lot of noise about it, and then both of them got burned eventually.

Well, I mean, I guess that would make sense, if the attribution that's sort of commonly agreed upon here in the West is that if 29 is SVR, the foreign intelligence service, and 28 is GRU, the military intelligence, then it would make sense that one would be more focused on information gathering and the other more on the active measures you mentioned.

At least to the extent that researching the Dukes, of course, the question of attribution was of interest. Primarily because in some situations it helps understand the motivations of the attacker and it helps assess how they may evolve and who they may be targeting in the future. And so for us what was meaningful was that the Dukes very consistently targeted ministries of foreign affairs, they targeted think tanks, especially foreign policy-related think tanks. They targeted embassies, for instance. So they very much seemed to be focused on foreign intelligence gathering kinds of targets.

And that was also one of the reasons why we were back then already quite confident that – we don't know whether the people with their hands on the keyboard were employed by a federal government or not. We don't know if they work for an intelligence agency or not. But we were confident, and remain confident, that the ultimate benefactor of that activity was a government. Because no one other than governments gets significant value out of that type of foreign policy information, and that type of foreign policy espionage, essentially. So that's why we were quite sure already back then that it has to be sponsored, at least, by a government.

Okay. Now since MITRE chose this group as an example for this year's testing, is there anything that stands out about Dukes as an APT group? Were they first for something, anything that makes them stand out from the mass?

There were a couple of interesting aspects to them that I think are still very much applicable. One of the interesting things was the type of attacks they did. They primarily had two kinds of attacks, at least the majority of the activity.

So they'd do smash-and-grab type of breaches. For instance, for a few years, twice a year, from late January to late February and again from late July to late August, they'd send out thousands of phishing emails. Not tens or hundreds, but thousands. They still picked targets relatively specifically, so most of them were relevant for their goals, but they'd send out the

same phishing email to a couple of thousand at least. And those phishing emails were quite generic, quite common. They used eFax themes, for instance, which was a popular theme also for ransomware at the time.

And then they'd seemingly wait and see how many fell for that. And if out of a couple of thousand they got some tens, for instance, what they'd do was they'd rapidly deploy one of their tool sets on those systems and grab as much data as they could. And they'd expect to get burned at some point. And then they'd go back and sift through the data and figure out which targets were interesting enough to actually target again, or go after again. And then they'd go back and use a different tool set and go in much slower, much stealthier, and much more targeted, and kind of go back and then lay low and gather information.

In other situations, from the smash-and-grab, they'd actually try to quickly switch to the stealthy approach, so they'd quickly switch tool sets, they'd drop other malware on the same systems, delete the original one, or they'd quickly move laterally to other systems, and on those other systems they'd use completely different sets of tools, and they'd kind of try and switch on the fly from making a lot of noise, as long as they'd get as much data as possible, to trying to staying there stealthy and persistent for a long time. So that was definitely interesting.

Well, that fits my perception of what these groups are about. So was Dukes the first to do that, or is that something that they brought into the common knowledge of the other APT groups now?

I wouldn't go as far as to necessarily say they were the first. Probably they weren't. But at the time there weren't many publicly known groups employing that type of a strategy, that type of a methodology.

Okay. So what do you think is the future of the Dukes?

I think the Dukes as such have already ceased to exist. Sure, there's been some clusters of activity since, but I think these days the Dukes are a useful but already historical case study. I think the people in the organizations behind the Dukes have probably long ago moved on. I mean, the people, the skills, the know-how, none of that has disappeared, it won't disappear, but some of the teams may have been disbanded or merged into other teams or reassigned, or who knows what.

So I'm confident that the legacy of the Dukes lives on. But I think from a defender's perspective, it's no longer meaningful to try and fit new activity into the same old mold of the Dukes. If the same people are still at it, they've learned enough to switch their tools and switch their tactics at this point.

So for us as defenders, it's still a useful case study, but not something you're likely to encounter live out there.

I don’t think you’ll find the exact same tools out there anymore, or all of the same techniques. But certainly many of the techniques they employed are still useful, still applicable.

A lot of the finesse of the Dukes as well was, for instance, they seemed to always have a good understanding of the human side of it. And what’s hard for organizations. So they knew when to move fast. They’d do the big phishing campaigns with phishing emails that looked like they were ransomware or whatever spam, perhaps with the idea being that it will probably be noticed because it looks like spam, but it also won’t be investigated very quickly, because it doesn’t look like anything interesting or targeted, so it won’t be a high priority. And so they’ll have time, a couple of days maybe, to go and grab as much as they can. They knew when to move slowly.

They understood how the incident response may proceed and how to make it harder. They had some really interesting ideas on how to try and hide and maintain persistence. So a lot of the tricks they did, and especially the understanding of how humans and the defenders work and what their weaknesses are, that’s something that is still applicable.

All right. Hey, thank you for taking us down memory lane, and taking this retrospective of your research from half a decade ago. Thanks, Artturi.

My pleasure.

That was our show for today. I hope you enjoyed it. Make sure you subscribe to the podcast, and you can reach us with questions and comments on Twitter @CyberSauna. Thanks for listening.

Categories

Podcasts, Threats & Research