

## Visualization of Network, Web Application, and Database Data – an Interview with Raffy Marty – author and Chief Security Strategist for Splunk

Listen to the Podcast [here](#).

**Brian Contos:** Welcome to the Imperva Security podcast. I'm Brian Contos, chief security strategist for Imperva. If you would like to learn more about this subject and Imperva, visit [Imperva.com](http://Imperva.com), check out our [blog](#), follow us on [Twitter](#), or send us an e-mail at [blog@imperva.com](mailto:blog@imperva.com).

Joining me today is Raffael Marty. As Chief Security Strategist for Splunk, Raffy is Customer Advocate and Guardian, expert on all things security and log analysis.

Starting with IBM Research and Price Waterhouse Coopers' consulting, then ArcSight and Splunk, Raffy has been in the log management and analysis world for many years. He has built numerous log analysis systems, and implemented use-cases for hundreds of customers that deal with log manager challenges on a daily basis.

Currently, he uses his skills in data visualization, compliance, security metrics, and risk management, to solve problems and create solutions for Splunk customers.

Fully immersed in industry initiatives, standards, efforts, and activities, Raffy lives and breathes security and visualization. His passion for visualization is evident in the many presentations he gives at conferences around the world, and his book, "Applied Security Visualization". In addition, Raffy is author of "After Glow", founder of the Security Visualization Portal, and contributing author to a number of books on security and visualization.

Welcome to the show, Raffy!

**Raffael Marty:** Thanks a lot, Brian.

**Brian Contos:** So Raffy, it's been about, I guess, a month since we ran into each other at the IANS conference in DC. What are you working on these days? What's new? I know you had a lot of interesting projects under your belt.

**Raffael Marty:** Well, my day job is still at Splunk, where I'm working as Chief Security Strategist. I'm working a lot of strategy about different markets, what markets are we going into.

I'm also planning for a new platform for the dome, that's coming out mid-year. Building a lot of apps for that.

But the really cool stuff, I guess, I'm still doing in my spare time as usual - I guess it's the same for everyone.

I'm still doing a lot of research on visualizing security data - I just co-wrote a paper on insider threat visualization.

## Visualization of Network, Application, and Database Data with Raffy Marty

I'm finding myself looking at a lot of tools lately - just visualization tools, and different technologies. I've seen a lot really great things around JavaScript, and visualizing data with JavaScript. I've looked a lot into the Flash side of the world, on how to visualize data. I'm doing all kinds of stuff.

I'm looking at a lot of papers for conferences still, to see what's going on there. I'm on different program committees. It's usually interesting to see what the new research is that's out there. So I'm definitely keeping busy right now.

**Brian Contos:** Yeah, some of the work that you've done around visualization's really pretty incredible.

To some of our listeners out there that might not know who you are, or of the book you wrote on the subject, maybe you could give us a little bit of background - what the book was, or what the book is, I should say. And what it's about, and where they might find it.

**Raffael Marty:** Sure. Well, the book is called "Applied Security Visualization," it was published by Addison Wesley last August. The book is basically an attempt of mine to kind of do a couple of things.

One is: I got really frustrated when... I have been doing log analysis for the last, I don't know, ten years, and it is always textual. You look at these textual log files, you read through them, and at some point you have massive amounts you need to analyze. And I realized, text is just not the solution to this.

And I started playing around - at my former employer, at that point. We built a feature in the product that let me generate link graphs, so that I could see a relationship between different properties and events. And I realized: there is much more to this visualization field than a lot of people realize, and I started playing around it.

I started actually, also, writing my own visualization tools - together with a co-worker of mine, Christian - back in the day, which we called After Glow. And that tool is basically a very simple script that takes some input data and generates link graphs that can show you relationships between different data sets, or different entities in the data. And it was such a simple tool, but I got a lot of traction - a lot of people liked it, they started using it for different use-cases, and security for visualizing net flow logs, for example, or firewall logs.

And so I just kind of took that further and further, and at some point I realized: hey, I have enough content that I can write a book. So I pitched it to the publisher.

The goal of the book is really to show people, hands-on, here is how you do some of these things. Here is how you utilize visualization for your own data, and here is how you... the tools out there, here's some of the mechanisms.

And maybe one last point about the book is that I think there is a big gap between the security world, where we have people that understand networking and routing and operating systems, and then the visualization world.

I call this the dichotomy of security visualization - where the visualization people, they're really good at perception, and depth queue theory, and code theory, and all these things. They understand how people interpret graphs. They know how to build really nice graphs, and what types of graphs to use for what purpose. But they don't understand the security realm - they don't understand our jargon. They don't understand how the Internet works.

## Visualization of Network, Application, and Database Data with Raffy Marty

So I'm trying to kind of bridge this gap, a little bit - to bring these two worlds closer together, and that's part of what the book tries to do. To give an introduction to visualization, the most important things you need to know in order to make good graphs, and bring these two worlds together.

And always in a very hands on way, that you can actually go off and do this - I don't like books that are just about theory. I need to have concrete examples. And I hope I was able to put that into the book, and give people something to work with.

**Brian Contos:** Raffy, besides having lots of cool graphs to look at, what are, really, the advantages of visualization?

**Raffael Marty:** Well, one, is that you already said it - sometimes a graph is just inspirational art. And I actually - it's not even funny - a lot of times when I generate some of these graphs, they inspire me to go down a certain path. I might look at a certain firewall data set and generate a certain visualization, and it just inspires me to figure out: hey, what else is in this data set, what other things could I do? So, it's definitely something that can be a fun way of playing with.

But for the more serious, or the more tangible result - if you have large data sets, it's just really, really, time-consuming and resource intensive to go through all the log lines, read through them, and understand what they're about. If you have a log of 10,000 lines that you start reading, by the time you hit number 100, you probably forgot what's in number 1. And you can't really put that picture together, of what's really going on.

So, the visualization, or some graph, or image, can oftentimes help you see certain patterns in this data. And this is definitely one of the big benefits. You see patterns right away. You can see: hey, this one user here had the most activity, much more than the other users. So it helps you find patterns, clusters, outliers.

And what I usually say is, visualization... for some of the use-cases I used it for, is one: to explore and discover. So I have a certain data set, I generate a certain visualization, and I can use that to explore what the data is about. And I might discover certain things.

And then, a second use is: I have certain questions about a data set. I want to know whether I had any attacks on my network that were successful - so I might try to visualize that in a certain way, and try to answer that question. Hopefully I can answer it, with the right visualization.

And a lot of times what happens is, I might answer this question, but new questions come up that I didn't even anticipate. So suddenly I realize - what is this? What is the cluster of events here? And start investigating that. So, new questions come up.

I can also use visualization to communicate information. That's probably something most of us use charts, or graphs for - if you generate a graph in Excel, for example. You can use that, and exchange the graph, or the information, with someone else, and they can understand it much better.

There's a multitude of uses that visualization really brings to the table, and a lot times it's really just speeding things up, and giving you a tool that lets you understand your data much better.

**Brian Contos:** So, you've applied visualization to a lot of areas - not just IT, or not just security. Taking a broader perspective - from what you've learned from writing the book, as

## Visualization of Network, Application, and Database Data with Raffy Marty

well as creating the software tool After Glow - what's some of the really interesting information that can be derived from visualization?

**Raffael Marty:** It's various things. Oftentimes, if you just generate a picture from your data, it's interesting what comes up. Because, let's say you're coming to me and you're giving me a firewall log file. And you say: hey, Raffy, visualize this firewall log. I might understand the log files really well, and I might be able to generate for you whatever graph you want - but in the end, I can generate you something, and you're going to be like: yeah, now what?

So, I think it's very important that people understand, hey, I need to really define what I want to get out of it. If I just generate some kind of report, I might just be lost in the end. But if you give me a use-case, I can start working with you, and say, OK, if you want to know what the distribution of the attacks was, based on geographical location, then we can work with that and I can generate a graph that hopefully communicate the property really well.

This is just something I'm trying to do with the book, or I'm trying to do with it, is to identify use cases that people have with certain data sources. For final logs, these are the things you want to do for intrusion detection logs here some things you want to do, like signature tuning for example.

Database logs, you want to know whether there are any funny or malicious activities hidden in your database logs or are there any processes running on the database they shouldn't be running. In compliance, you want to have separation of duties. There's a slew of use cases and probably in every market segment, any area that has electronic locks or audit data, you can identify use cases that would really benefit from visualization.

**Brian Contos:** Could you share maybe one or two real-life examples, as they relate to you, either network devices and/or traditional network security devices?

**Raffael Marty:** Sure, what I find a lot of people doing, a lot of the Splunk customers also there, they are looking at net-flow data. You have traffic flows, you see who is communicating with who on the network. These are just humongous amounts of data. A lot of times, what I want to know if I'm going into an environment and I need to assess what's here, I want to know what machines are talking on the network and what ports are they talking on. What services do they offer?

The easiest way to do this is not going into the log files and starting reading all the net flows, but if I generate a link-graph, that has nodes identifying the machines, and I connect them based on their communication, then I can see - kind of like a network topology view, who is talking to whom on these networks.

If I start overlaying more information on that, for example, I color the edges by port numbers, or bunches of services that they offer, then I can very quickly identify, "oh, these are machines that are using the weblog or these are mail servers because they are accepting connections on port 25, for example. Or these machines are being hit with SSH a lot."

So you can very, very quickly see what's the state-of-the-state, what's going on in the network, and you might realize already from that view that there are some funny things going on. For example, you might see that your DNS server is not just access on port 53, for DNS requests, but maybe also port 80, which also acts as a Web server, which is generally a violation of separating your roles in your environment.

## Visualization of Network, Application, and Database Data with Raffy Marty

So that's just one way of very quickly looking at something that's out there. Another interesting investigation that I did, was trying to find bot nets in a corporate network and this was inspired by a discussion of the Honeypots mailing list a few years back, where the premise was that if you had an infected machine, this machine will get an update at some point from the controller, basically a new code.

When that update comes in, it's a certain size. The machine is being updated, will probably propagate this update to other machines and these updates are going to approximately the same size as the packets it came in, but it's not the same. If it was the same, I wouldn't have a problem, but just write a correlation rule or something in Splunk or in some log management tool.

But because the sizes are slightly different, I need to use visualization to see the pattern around the size - or just clusters around different sizes. I was able to use parallel coordinates, which is a more advanced visualization tool to look at the traffic, so I can figure out here, I have a cluster of activity and I was able to find the bot nets, and the affected machines in my network.

**Brian Contos:** Let's take that same idea in these examples, and see how can we apply this to applications and databases?

**Raffael Marty:** I actually wrote a section in my book about database audit log visualizations. There's two big problems in databases. One is, if you turn on auditing, you usually have vast amounts of logs you need to look at. You guys are in that business. The second problem is, and this is something I really like in Imperva for, if you actually turn on all the auditing on the database, you bring database performance down.

Especially if you start looking at all the activity, if you look at all the selects and all the updates on the database, the database can't handle it. This is really what I like about Imperva, you basically put your appliance in line and it analyzes the traffic, and it can record all the database activity without impacting the database itself.

Now I can really look at all of the audit records and again, I think it's interesting to look at visualizing all of this because it's so much data. One of the things I've been doing is, I basically looked at what users are accessing what tables and what activity are they executing on the tables. You might find people that just read data from the tables, you might realize that some people are reading and writing them, or updating tables and inserting different rows.

Then you have a whole set of administrative activity on the database also, where you alter tables, where you drop tables or where you create new table spaces, or people execute maintenance commands on the database, like they check for free disk space, or they might change certain properties in the database.

You can look at all of this together, and now if you see that someone, that's really in the finance group, for example, usually reads from the finance tables, and maybe updates them, but suddenly he executes some database command and you will see individual output very quickly, that there is a cluster of users around this finance group that update a certain number of tables and then this one user will be kind of on the side of this cluster and he will have a connection to some of these other activities like looking at database setups or something.

## Visualization of Network, Application, and Database Data with Raffy Marty

Maybe he tried to execute certain commands from the database that he shouldn't be. This will be very quickly visible, inside of this big cluster of normal activity that will be an outlier very obvious that this guy's up to something no good.

You can take this to any degree of monitoring different users. You can correlate the database users with the operating system users, if you have all of that information, you can correlate that back to the groups they belong to and again build these kind of clusters of group behavior.

This is probably the biggest benefit of visualizing database logs, you can look at groups of similar behavior and see what they're up to. You will very quickly, as I said, see these outliers happening in the activities.

**Brian Contos:** One of the areas, with Oracle, for example, is if you have auditing turned on and, to your point, it just has a performance impact, it will actually only track well-formed SQL queries. So if a situation occurs where there is a malformed SQL query, like we see in a lot of reconnaissance events and certain types of attacks, it just simply won't show up in that audit log. Not to necessarily plug Imperva, but, by using a solution like Imperva, one of the things we do is, we get well formed and malformed, so you have a greater amount of information to which you could then leverage for visualization.

**Raffael Marty:** Absolutely, I think that it's crucial - sorry to interrupt you here - but I think it's really crucial that you have all of these log records, because otherwise you're going to miss half the information and you don't even realize that someone is trying to break out of their usual role behavior. You wouldn't even see that outlier.

I think that's a very important point for visualization, that the data that comes in, or the visualization is only as good as the data you get into it.

**Brian Contos:** I would even say that, because of the visualization, you could actually deal with much larger data sets, more easily. In some cases they might say, "I don't want a log as much, because I'll never be able to get through all the data." But with visualization, that becomes moot.

**Raffael Marty:** Absolutely.

**Brian Contos:** Well, Raffy, we have time for some closing thoughts from you.

**Raffael Marty:** Sure, what I'm trying to do is really building a community around this whole security visualization topic. I started a few years back a website called [secbiz.org](http://secbiz.org), it's a portal where people can discuss security visualization and people can submit their graphs they have, use cases they came up with, tools they built.

We have a pretty active community up there, but were always looking for new people submitting and helping discuss and questioning what we do or ingesting new things. So anyone who wants to drop in on the website is very welcome too. As of two weeks ago, we finally have a mailing list also. There is a twitter stream @secbiz.

I really want to grow this community and get more people to start realizing that there is a lot of benefits in visualizing security logs.

**Brian Contos:** Well, great stuff, Raffy. Thanks so much for joining us on the show today.

**Raffael Marty:** Thank you very much, Brian.

## Visualization of Network, Application, and Database Data with Raffy Marty

**Brian Contos:** If you would like to learn more about this subject and Imperva, visit [Imperva.com](http://Imperva.com), check out our [blog](#), follow us on [Twitter](#), or send us an e-mail at [blog@imperva.com](mailto:blog@imperva.com).



North America Headquarters  
3400 Bridge Parkway  
Suite 101  
Redwood Shores, CA 94065  
Tel: +1-650-345-9000  
Fax: +1-650-345-9004

International Headquarters  
125 Menachem Begin Street  
Tel Aviv 67010  
Israel  
Tel: +972-3-684-0100  
Fax: +972-3-684-0200