EMA Perspective:

SIEM 2.0 – Filling the Visibility Gap and Making Log Data Useful

Scott Crawford, Managing Research Director for Security & Risk

Formerly Chief Information Security Officer for the Comprehensive Nuclear Test Ban Treaty Organization & its International Data Center in Vienna, Austria

I am Scott Crawford, Managing Research Director for Security and Risk for Enterprise Management Associates. Enterprise Management Associates is an industry analyst firm focused on IT Management and all its aspects - IT management software, services and practices. And that includes security management as well as systems application and information management and business intelligence.

My personal background: I became an analyst after having been the Chief Information Security Officer for the Comprehensive Nuclear Test Ban Treaty Organization at its International Data Center in Vienna, Austria and that bleeding edge experience if you will, is what gave me the motivation to push ahead into what is coming in security management and to help my peers get a better handle on where they need to go in order to improve the approach to security management in the enterprise.

Security in the "Age of Big Data"

So, I have been talking and writing a lot lately about the challenges of managing security in the era of big data which we are fully in now and enterprises as well as even small businesses realize this just about everywhere. We certainly saw evidence of that with the Wiki Leaks case in the last few months given the sheer volume of information that governments have to deal with for example and how it had an impact on data management. And to just give you some idea of scope what we are talking about here in terms of big data Eric Schmidt, CEO of Google, speaking at a conference this past summer mentioned that between the dawn of civilization and 2003, all of five exobytes[PH] bytes of data (it is a billion, billion bytes of data) have been created over that entire time. Now we are creating that much data every two days. So it is just, it is really hard to get your hands around that concept of what that really means for businesses of all kinds is that first of all, the sheer volume of sensitive information that we as security professionals have to protect has really mushroomed.

At the same time, threats and vulnerabilities have also mushroomed as attackers have become more sophisticated, as they have followed the money if you will, and realize that sensitive information is directly linked to tangible assets as well as the things like you know, economic advantage, competitive advantage in terms of corporate espionage, things of that sort. So this has really raised the stakes on information and information security management but has also just threatened to drown organizations not just in primary data but also in the monitoring and log data that they collect which is a really substantial part of the total volume of information handled in any particular organization. Now, data management has already posed real serious challenges for security professionals who have been at risk of drowning in data for as long as I have been involved in the business. But the way things are going with the trends toward just the sheer explosion in data and the volume of monitoring data and the number of things that we are looking for in security, it really does threaten to drown organizations and they really need tools to get a more effective handle on this if they are to keep up with a truly competent security strategy at all.

So, one of the questions I am asked in looking at the challenges of big data and security management is, "Are there any opportunities that these changes pose?" and yes, I believe there are; I mean for thing, we have access to a lot more data that gives us a lot more insight into the real nature of the actual threats we face, our real risk posture then we ever have before just by virtue of the number of things that we are collecting, the number of events that we're collecting, but we have to learn how to use this information a lot more intelligently and nothing highlighted that more than or has highlighted that more than the Verizon Business Data Breach Investigations Reports. In the first of those reports released in 2008, they found that 82 percent of the victim organizations in these cases had actually collected data that would have informed them about the threat they were dealing with, the threat that was afforded their environment but this information was either ignored, not noticed, not acted on. That figure actually increased in the 2010 report to 86 percent of the cases covered. So, it is not that we are not collecting the data; it is that we need to learn how to use it more effectively.

We need the tools to give us access to this data in a meaningful way and to enable it to enable us to give us the tools that we need to have better insight into what we are actually dealing with and in my view as we get into this era of big data in both the primary data we protect as well as the monitoring data that we collect, I think we really need to take a page from the world of Business Intelligence (BI). We need to be developing tools that make access to data more actionable, more flexible particularly because in security we may not know from one day to the next what we are dealing with. We do know some things as far as you know, the nature of the data that we are trying to protect, the ways that it could potentially leave the organization but we don't know exactly how that's being done in all cases. We don't know things like emerging zero day of vulnerabilities for example that are being exploited by attackers doing their own research. So, these tools need to be flexible, they need to be very adaptable, very swiftly adaptable, they need to be very real time in the way that they can apply this intelligence and they need to make data accessible in ways that make it easy for individuals to put information together, to build a sequence of events, to deliver context, and deliver better insight into actually what we are dealing with in terms of threats and risk management.

So how could an enterprise security log and event management system help address some of these problems? Well, first thing is that these types of solutions are designed to collect and particularly when it comes to log management, to collect monitoring information from all types of systems throughout the environment and not necessarily even just IT. It is important to collect, to really have context into what you are dealing with, it is important to collect not only from all different types of IT monitoring systems, not just security management systems or security defenses but also things like network management, systems management, application management, application integration tools considering that attackers are looking for the weak points in applications particularly these days, but also things like physical access, badge reader access that tells you when an individual has physical access to a certain part of the organization or when physical access is attempted or looks anomalous, someone using that individual's badge to gain access at a time that stands out because it does not look typical.

These are all the kinds of things that can be put together in a truly comprehensive solution and that's really what you need; to get the context into the really difficult to identify, the types of threats, the types of risk exposures that can be very subtle, which attackers really value because it makes it easier for them to hide their activity. Attackers, external attackers as well as internal personnel that are acting outside, that are authorized privileges you really need that breadth of context so, having a log management solution that aggregates logs and event monitoring data from all types of environments, from all types of solutions, really much more important now than it ever has been before.

Discovering what's really happening via SIEM

In the past a lot of solutions basically just called events, of all types of monitoring solutions. If you recognize an event, it was elevated up to a centralized solution but it may not have had the real context that you need to really drill down into the real nature of what you have to deal with. So this aggregation of data speaks of the challenges yes, of security in the era of big data but it also speaks to the opportunity that we have if we have the right tools to make use of it.

So, part of what that tool set also needs to include is the ability to make sense of this data, rationalizing and normalizing this event data, really important so that you know that you are comparing apples to apples when you are looking at events that are reported in a dissimilar way. But also so you don't have to be you know, very expert and very arcane terminology, taxonomy, syntax of any one particular solution; again to make it actionable across the board, across the entire spectrum of monitoring solutions that you are collecting this event data from. They also need the ability to reduce this data into more bite-sized elements if you will, to be able to build insight from a collection of building blocks if you will, in order to be able to build context around specific events and to develop that context which may you know, may develop over time. You may need to collect events over a long series of time to really get the real context of what you are dealing with. So the ability to put together this insight from atomic or elemental level, bits of information if you will, to build this sort of insight, to build this context over time, that building block approach is really very important. Of course, you have to have tools for reporting particularly important in this very compliance sensitive world in which we live today, as is the ability to collect, manage, and maintain this log and event data over long periods of time which may also be required to meet regulatory compliance.

How Log Data Enables Proactive Security

So why do I emphasize this need to collect all this volume of data across all these monitoring points across all this long period of time. It is very simple. Attackers, the militias, those who don't have the best interest of the organization and its customers and its stake holders in mind are getting very, very subtle; very, very precise in the ways that they target risk exposures. So things that might look like a significant event to certain type of a monitoring system in one type of a silo like say network management or systems management or an intrusion detection system might not look like anything particularly significant, it might look like something you have seen hundreds or thousands of, it might easily get lost in the noise. But, if you are collecting this breath of data across all these silos, then you can start picking up evidence of the sequence of actions, the sequence of events that leads to a real exposure that leads to a real incident. The extent to which you can collect this information and be able to act on it, as close to real time or better yet as ahead of these events as possible makes it much more possible for security organizations to move from much more reactive approach to security where the team is reacting to an incident after the fact or to vulnerability after it has been disclosed to having more insight into where the exposures really are and being able to become more proactive in building a better approach to defending the organization against these types of risks.

SIEM & BI Parallels

So, what are some of the additional benefits of security log and event management, systems log management, security event management systems that people don't normally think of when they think of this space because they think of security event, log management as being looking at security events from security point products: firewalls, intrusion detection prevention. What we are not thinking of in my view is the bigger picture. To go back to what I was saying earlier about business intelligence, I think we really need to start thinking more in the way that the BI community does. We have this huge volume of data, we have all kinds of information we can mine from it, but what are the techniques that will be effective for us in this space.

I think the first place to start is to consider, "What are the foundations of BI itself?" What are the things that really made business intelligence fly as a science in its own right is the foundation it is built on, the relational database for example. The relational database made it very straight forward to derive insight from large volumes of data because it made putting together these atomic bits of data in a way that made sense using natural language, something is part of, look for something in a set of something else and just give me the results. We historically haven’t had very much of that in security; we haven't had the same sort of building block approach to deriving insight from these large volumes of data that we have to deal with. So taking that sort of approach is the starting point, I think it is some place that we need to begin. We also need to begin to make what we realize from this data actionable in new ways like the ability to visualize it more effectively, definitely the ability to do this very flexibly, very fluidly and on demand. Again going back to the parallel with BI, relational databases, you can build a query in SQL very straight forwardly, very quickly. We need that same sort of capability in security because the landscape is constantly changing.

Security is fundamentally gamesmanship. When we build a better defense, the attacker is going to be looking for new and different ways to exploit what we have or more effective ways to exploit what we have that heretofore have not been exploited successfully; a way to do so with greater stealth and less obvious, less visible attack methods. We need to be able to adapt to those kinds of things with the agility that the militias have. Our tools need to be able to deliver that kind of insight from the information that we do collect with that same level of agility.

Many Audiences for Log & Event Data

Another way that we really need to expand our thinking about how we deal with all this log and monitoring information is that this isn’t just about security and for that matter IT operations monitoring data isn't just about IT ops either. A lot of the value that we derive in security management comes from operational management tools and operational management data and vice versa. All this data that we have, we can apply a common set of tools to derive insight into things not just about our risk exposure, not just about the militias, not just about the actions, the behaviors, the threat landscape itself and how it is affecting our organizations but also about things like you know, performance, availability, resource integrity, are our complex information systems doing what we want and expect them to do? With a common set of flexible tools for building this very rich and detailed insight over time, from this volume of data, making that more actionable, more realistic and more available to us in the here and now can do a lot to really help organizations get a leg-up on all kinds of aspects of management. We really need this universally across the board, not just in IT. So again, making the ability to derive insight from monitoring data more actionable, more flexible, more easy to use, so you don't have to have a PhD in this stuff just to build a syntax for a particular type of a query will make security management more effective and give us the tools that we need to take advantage of the volume of data that we are already collecting.

SIEM 2.0 – Filling the Gaps

So, how do today's emerging approaches differ from what some people might think of as a SIEM 1.0, security event management, log management 1.0 type approach? Well, 1.0 solutions really were kind of a one-off if you will. We had a lot of security management tools that generated a ton of data, we needed to make some sense out of this volume of data we were collecting from all these tools which meant that we began to build and use tools that would cull these events, clip them off the top of the waves, if you will, that would give us the most significant events that we felt were the highest priorities to us as security professionals which meant that a lot of times, they were very complex systems because of what they had to integrate and rationalize. They were very inflexible because of the nature of the rules that had to be put together to recognize specific types of events. They didn’t necessarily lend themselves to being all that easy to use.

A lot of times they used very arcane scripting languages that you had to be a real expert in to know how to use effectively and they were you know big and heavy to ploy; I mean, these were complex applications in their own right requiring a fair amount of compute, an awful lot of integration resources which often times meant a lot of services to be brought in to integrate all these pieces together, a lot of storage, certainly still quite a bit of storage for collecting the event data that we have today but the complexity of putting first generation systems together was really overhauling for a lot of organizations which meant that a lot of businesses just weren't able to have access to the level of expertise that they needed in this area.

Today what we see is a lot more push towards tools that are a lot more flexible, a lot easier to deploy, lot more intuitive in how they make the ability to analyze information and gain insight from this information more directly actionable if you will, by making them a lot easier, a lot more natural to use, so you don't have to have a PhD in these things in order to know how to derive a great deal of use from them and because of this flexibility and the breath of what they can collect, particularly if they are collecting log data across the board, that makes them a lot more useful for building a more, a richer sense of context over a longer period of time from a lot more data we had before and that's exactly what we need in this era of exploding data.

Something else about emerging solutions that are different from first generation solutions is a lot of time first generation solutions were either they came from the event management world or they came from the log management world where you are centralizing and aggregating all this log data. You may or may not be applying you know, culling events from it but in an event management systems you are culling events from all this data. Reporting may have been even you know separate still in terms of non-real time event management for reporting. So there is these isolated segments with log management here, event management here, reporting in some cases or maybe even segmented in some cases. Today what we are seeing is a lot of these coming more together because security professionals know that they need access to the underlying log information, they need the events, they need to know when they need to be aware of things in their environment but they also know they need to drill down, they need to get to the source data in order to really build that context, and also in order to build more effective event rules in the future, they need to build these rules from a wider breath of information which is again driving these solutions more together in emerging generation tools.

SIEM Buying Considerations

So what to think about when you are considering a log or security event management solution? Well, one of the first things you need to think about is does it really adapt to your organization rather than forcing it the other way around, forcing your organization to adapt to the solution. Is the solution sufficiently flexible to adapt to your deployment requirements? You may have a requirement for you know centralized deployment where you are just collecting from one central location or you may be a very distributed organization or you may be collecting log data from a number of peer or subsidiary or partner organizations that may be globally distributed. What is the architectural requirement in your organization and does the solution that you are thinking of meet those requirements for your organization? Is it as effective in a large enterprise as it would be in you know, a smaller organization where keep in mind that even within a large enterprise, you may have individual organizations that have their own monitoring and reporting requirements? These requirements need to be met as well too and yet at the same time the organization as a whole may need to collect all this data and more across the organization in its entirety.

So is your chosen solution sufficiently flexible to adapt to these different types of deployment requirements. It has to be agnostic; it has to be able to incorporate information from a wide range of tools, not just a wide range of known security tools but management tools of all kinds, all kinds of logging data, bear in mind that system logging data is not created equally across different types of environments. Data that people will be familiar with for the Syslog environment can be very different from what's collected in a Windows environment, the way that it’s handled, parsed and used can be very, very different. Even more different still, different types of application systems and the components that make them work, physical monitoring, physical access controls. All these things will factor into your security strategy or should and the data that they collect should factor into your log and even management strategy as well.

LogRhythm’s AI Engine – Correlation that works

So what do I think has really held back first generation SEIM 1.0 type approaches? Well, first of all, it’s just been the complexity of these solutions as much as anything else that can be really difficult, even for a sophisticated security organization to get their hands around these things because they are so complex the challenges of integrating with so many different things has made them that way over time. So there’s a real need to reduce that complexity in their use so you can gain real intelligence out of them. They’ve also been real brittle, if you will. They’ve been fairly rigid in terms of the way that rules are implemented and changing those flexibly in order to meet real time demands in the fast-changing threat landscape has not always been real straight forward for some of these solutions. So what I like about the LogRhythm approach is that LogRhythm has always had this notion of a building block approach to taking simple elements of the information that you have and building really sophisticated queries and rules from your data in order to let you know what you really want to look for and making it a lot more accessible to ordinary people.

This gets back to what I was saying earlier about we need to take a page in security from the world of business intelligence and data management where the relational database really revolutionized our approach to data management. It made natural language the tool for putting together these queries, these inquiries into large volumes of data and making it very actionable to ordinary people. LogRhythm has always had a very similar approach to security information and event data. And so what we see LogRhythm doing today is really elaborating on that and making the development of these really sophisticated rules even more simple, even more straight forward with a gooey [ph] based approach, a visual based approach for building these elemental rule blocks that can be created and purposed for any sort of analysis that would be useful.

You can define specific things, like a threshold, so much information that you see moving across a certain point within a certain period of time. You can apply that to almost any type of exploit where something like exfiltration [ph] of data would be of interest. You can create that rule block, use it over and over again. So simplicity in creating analytic tools, the reusability of those tools through things like rule blocks, the gooey element of the AI engine that makes putting these together a lot more intuitive and a lot more straight forward, coupled with the volume of log data from all these different types of resources across the entire environment places this breadth of information at your disposal and makes it a lot more actionable to really derive really meaningful insight from.

Correlating ALL logs is Critical

One of the things that I think has been a problem with earlier generations of solutions, SIEM 1.0 type solutions, is as I mentioned earlier this inflexibility in terms of being able to modify rules or having rules that are so complicated to modify or having to be such an expert in how to modify them that makes it very impractical for a lot of organizations to be able to adapt them on demand. But there’s a real need for this as I alluded earlier that with a fast changing threat landscape you really do need to be able to adapt your monitoring and event solutions to the realities of this landscape as it’s changing.

Let me give you an example of how this might work out in practice. You probably have an idea of the nature of impact of certain types of threats. You would be able to recognize the exfiltration of data or large amounts of data at a certain period of time or within a certain period of time. You would probably recognize certain aspects of anomalous access to sensitive information resources that would tell you someone has insight into where to find that sort of sensitive data within your organization.

You can start there and work the problem backwards into what are the tactics that were actually used to get there. To do that you are going to need to adapt the correlation of event data to give you that insight so that you can work the problem backwards, if you will, and refine your analysis to get to the point where you really do recognize what's going on in your organization. To do that you need to be able to adapt your correlation rules on demand, in real time if you're dealing with a threat that's currently afoot or something that's emerging right now.

LogRhythms AI Engine makes that a lot more accessible to a lot wider range of ordinary people. People who are expert at security, yes, don't necessarily have to be an expert in the arcane details of specific types of rule sets or the way that rules are constructed in order to get to this data that's more actionable, that's more useful to them in the here and now then a lot of historic solutions have really made available.

Detecting Sophisticated Threats with SIEM

So let me give you an example where having access to this breadth of log data coupled with an advanced correlation engine to really derive meaningful insight from all this data can be really useful. Let's take a fairly straight forward example where you're collecting monitoring data from a badge reader system and you note – the system notes that someone is physically accessing the physical location at a certain place at a certain point in time, simultaneously or it may be within say half an hour or so. That same individual appears to be accessing, or actually is accessing so far as the monitoring system is aware, from another continent within a half of an hour. Now obviously this is going to be an anomaly but there's a lot of systems that won't pick up on that because they don't have the full context of all the monitoring data from these very disparate systems.

With its advanced correlation capabilities LogRhythm does give you a way to not only pull this data from all these different resources but to incorporate it together in an advanced correlation through the AI Engine, coupled with things like time stamping where you have a universal time stamp that's applied to all the data collected from all the sources that will give you this picture in real time so that anomalies like this stand out more clearly. This is something that you might be using already for fraud detection. You may have a purpose built system for that sort of thing but you also have this information in the event data and the log data that you're already collecting so why not take advantage of it and use tools that deliver this insight for you from the information that you already have.

So, one of the most serious issues that organizations are dealing with today is the risk of compromised credentials. And just to give you an idea of why this is such a big issue, keep in mind that the more sophisticated attacker is going to want to get access to what the most sensitive users with the most sensitive access have in their organization. This could be consumer online banking accounts. It could be engineers with access to source code or intellectual property that differentiates a company in its competitive landscape. It could be the sort of thing where the individual has authorized access to things that are sensitive in a national security sense. But anomalous access, even on the part of an authorized individual, could pose a risk or could actually be a threat. These can be very, very difficult to tease out of event data if you don't have the breadth of context that you really need, particularly in the case where you have an individual with authorized access to sensitive government data. They're supposed to have access to that information but what they do with it poses a real threat to a government organization, say, or the national interest in an extreme case.

So you need to have visibility into more than just the individual's access privileges when an individual has access to their own accounts. You need to be able to determine when the way that they're using it does stand out. They're transferring a lot of data to a specific endpoint that looks like it stands out, it doesn't look right. The example I gave earlier of having one individual accessing a location physically at one point and then from another physical location within the same period of time, maybe a continent away, within the same half an hour, LogRhythms AI Engine gives organizations the ability to delve into this level of insight, into the log data that they have. Techniques like geolocation add to this by enhancing the ability to determine when this type of anomalous access really does stand out. Universal time stamping, another way to determine the sequence of activity that gives you greater context into these types of threats and helps make things like compromised credentials really stand out more clearly. When you having to work the problem backwards from the impact that you see to the way that these types of exposures are being exploited, that's the level of detail that you need to tease these subtleties out of the information that you're already collecting in your log data. So compromised credentials are one example of a use case where organizations may have historically been blind to a lot of the factors that they really need to know in order to determine when access is being compromised. But that's just one example.

There's a lot of other examples where organizations may be effectively blind to very real risks without the types of tools that they can have in LogRhythms AI Engine, for example. And another good example of that is privilege user monitoring. When high privileged individuals have very wide access or very sensitive access to information resources themselves, that access itself comes with some risk. I mean, there's no other way really to see that. You need some insight into this and there have been a number of cases where organizations have been exposed to these types of risks so you need some insight into how those privileges are actually being used in order to be aware of when you might actually have a real threat in your environment, otherwise you are going to be blind to it.

Another example of that might be early visibility into 0 day threats, which is what I was alluding to earlier. You have an idea of what the impact is and you might see evidence of the impact when you see data leaving the organization, when you see a sight compromised, when you see questionable activity in a database, but your security tools aren't necessarily picking up on these because they don't have the signature for it, they don't have the heuristic ability to recognize it. That right there is a pretty good indication that you may be subject to a 0 day threat that you're just not aware of yet.

So having the insight into the behavior that's led to the impact that you're seeing, you need that from your tools. You need the ability to apply what companies like LogRhythm give you with the AI Engine in order to be able to flexibly and easily analyze this data and to derive in real time what you need from this information that you're already collecting and you might be blind to it. Otherwise if you don't, if your rules are too fragile, too brittle, too rigid and inflexible, too difficult to modify to be able to see these things, you may be totally blind to them if you're depending on existing approaches or legacy tools.

The Evolution to "Data-Driven Security"

Data management security has become a number one priority for organizations of all sizes to the extent that I think in this year we're going to see a real emergence of approaches to what I consider data driven security. What I mean by that are approaches to security management that are directly dependent on data. We're going to see a direct analog to what we see in the scientific world where a lot of primary investigation has created these huge data sets but those data sets have a wealth of information in them. You don't need to repeat the experiment in order to glean new learnings, new understanding from the body of data that already exists in all the scientific data that we've accumulated.

Right here in Boulder we have a data place, if you will – The National Center for Atmospheric Research – which has accumulated a ton of data of all kinds from all different sciences. Now data scientists are beginning to apply their expertise to the data itself, creating this what Microsoft's Jim Gray once called The Fourth Paradigm, a new realm of scientific investigation that's centered on the data itself as the primary focus of investigation.

I believe we're on the verge of a turning point in security where security management is very data focused, data driven as I put it. But to really make that happen, we're really going to need to have tools that enable us to build these sophisticated queries, these sophisticated rules and analysis, but to build them in simple actionable ways. From very simple rule sets, from very simple rule blocks as LogRhythm uses the idea, in order to be able to build truly sophisticated insight into the volume of data that we have. We need these simple tools to deliver complex intelligence because we're not going to be able to get the insight we need from this data without them.