User-centred security event visualisation

Christopher Humphries

ANNÉE 2015

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Informatique

Ecole doctorale Matisse

présentée par

Christopher Humphries

préparée à l'unité de recherche CIDre
Confidentialité, Intégrité, Disponibilité et Répartition
Supélec / INRIA

Visualisation d'événements de sécurité centrée autour de l'utilisateur

---

User-centred security event visualisation

Thèse soutenue à Rennes le 8 décembre 2015

devant le jury composé de :

Hervé Debar
Télécom SudParis / rapporteur

Jean-Marc Robert
École de Technologie Supérieure / rapporteur

Isabelle Chrisment
Télécom Nancy / examinatrice

Christophe BIDAN
CentraleSupélec / directeur de thèse

Nicolas PRIGENT
CentraleSupélec / co-directeur de thèse

Frédéric Majorczyk
DGA-MI / co-directeur de thèse

I didn't go to university. Didn't even finish A-levels. But I have sympathy for those who did.

— Terry Pratchett

RÉSUMÉ EN FRANÇAIS

Alors que les systèmes d'information deviennent de plus en plus complexes, il n'est plus possible d'en maitriser tous les aspects et les mécanismes de sécurité traditionnels atteignent leurs limites. Les systèmes de détection d'intrusion ont été proposés pour faire face à ce nouveau défi. Ils offrent une approche automatique et rapide pour détecter et répondre aux intrusions en identifiant des motifs connus ou des situations qui divergent du comportement normal. Ils sont excellents pour répéter cette action. Cependant, de tels systèmes réagissent mal aux événements inconnus, déclenchent beaucoup de faux positifs et manquent des événements importants. Ainsi, il est aujourd'hui de plus en plus difficile de gérer les énormes quantités de données produites dans le cadre de la sécurité des systèmes d'information.

Pour cette raison, les opérateurs humains sont toujours nécessaires pour donner du sens aux événements de sécurité rapportés. Pour que leurs interventions soient utiles (en particulier quand la tâche est d'analyser des alertes), ces opérateurs ont besoin de comprendre des situations rapidement, d'obtenir facilement une vision globale et les réponses à des questions, ainsi que de consulter des grandes quantités de données contextuelles annexes.

La recherche en analyse des évènements de sécurité s'est récemment portée sur la visualisation. La visualisation pour la sécurité se place entre des solutions manuelles et automatiques, et vise à combiner le meilleur des deux pour proposer des outils efficaces en pratiques. En d'autres termes, des représentations visuelles permettent d'améliorer le processus de supervision et de fouille de données du point de vue de l'opérateur en améliorant la manière dont les données de sécurité lui sont communiquées.

La conception d'outils de visualisation dédiés à la sécurité est rendue difficile par plusieurs aspects : les formats de données et les protocoles sont variés, les standards sont nombreux et le matériel et les logiciels informatiques sont souvent paramétrés de manière très fine par les administrateurs pour répondre à leurs besoin spécifiques. Alors que certains domaines peuvent compter sur la nature immuable des problèmes ou la disponibilité de temps, les problèmes en sécurité sont trop souvent nouveaux et exigent en outre une attention immédiate.

⋈

Le processus de visualisation peut être vu comme une série de transformations de données. Chaque étape peut être spécifiée en utilisant une grammaire graphique, ce qui permet de construire une description concrète pour chaque visualisation.

À son niveau le plus élémentaire, une visualisation est un assemblage de composants graphiques, chacun ayant des attributs paramétrable. Ces variables visuelles peuvent être associées aux données pour obtenir une transcription graphique d'un sous-ensemble de ces données. La perception humaine étant hétérogène et changeante, chaque variable visuelle est plus ou moins adaptée pour transmettre les informations à l'utilisateur. Chacune d'entre elles doit donc être soigneusement sélectionnée en fonction des objectifs de la visualisation.

Lorsqu'on prend en considération l'expérience utilisateur, un nouveau processus d'interaction homme-machine émerge. Des boucles dans les étapes permettent des retours des utilisateurs et représentent un processus de fouille adopté pendant l'exploration de données. Pendant toutes ces interactions, les données graphiques doivent être retraduites mentalement par l'utilisateur. Pour alléger la charge cognitive induite, la progression de la fouille doit guider l'utilisateur et suivre une progression logique narrative proche de sa façon de penser.

⋈

Actuellement, la visualisation d'événements de sécurité suit une approche et des objectifs adoptés par des équipes chargés de la sécurité réactive des systèmes, et vise au moins un but parmi trois :

monitoring Dans le cadre de la surveillance des systèmes et des réseaux, les opérateurs utilisent des tableaux de bord de visualisations pour s'assurer que les métriques sont dans des plages optimales, dans le but de garantir la disponibilité des services et chercher des signes d'attaques connues et de comportements malveillants.

analyse Lorsque des anomalies ou des intrusions sont signalées par des systèmes de détection ou des opérateurs avec des outils de monitoring, des outils d'analyse visuelle sont utilisés pour mieux comprendre ces anomalies/intrusions. En explorant les données de sécurité, les analystes cherchent à trouver des explications pour ces incidents, à recréer les scénarios des attaques et à caractériser les faux négatifs.

reporting Une fois qu'assez d'informations ont été recueillis, elles ont souvent besoin d'être transmises au reste de l'équipe ou à des entités externes. Dans ces cas les rapports ont pour objectif d'informer des collègues, des cadres ou des externes dans le cas ou la gestion de la crise auprès des médias par exemple.

Pour accomplir chacun de ces objectifs, les outils vont proposer des visualisations adaptées. La surveillance favorisera des visualisations adaptées à la compréhension des données en temps réel et compatible à les données en évolution, donc plus simples et facile à appréhender. Puisque l'exploration de données est moins sujet aux contraintes du temps réel, les visualisation peuvent être plus complexes et configurable par l'utilisateur. Les rapports communiquant les résultats des deux étapes précédentes doivent condenser toutes les informations pour expliquer la situation ainsi qu'un contexte assez riche pour palier le manque d'interactivité.

⋈

La construction des outils de visualisation nécessite les connaissances de multiples domaines tels que les statistiques, la conception d'interfaces et la psychologie. Les experts en sécurité manquent souvent d'expérience dans ces domaines ce qui rend la fabrication d'outils ad hoc de visualisation difficile. Ils sont avant tout experts en sécurité, mais rarement entrainé pour la visualisation. Même si la visualisation sans entrainement fais parfois preuve d'innovation, elle peut aussi produire des résultats trompeurs.

Réciproquement, l'utilisation d'outils de visualisation pour la sécurité nécessite des connaissances en sécurité, en particulier pour analyser des formats de log et des configurations systèmes. Pour faire face à ces situations, les experts en sécurité se fient à leur ressenti pour comprendre la situation, identifier les problèmes, et trouver des solutions. Cette familiarisation se développe en instincts et habitudes qui peuvent finalement devenir de réels protocols.

Dans notre cas, les tâches à accomplir demandent l'exploration de multiples sources d'information différentes, ce qui implique le choix d'un outil spécifique à chaque fois qu'une nouvelle source d'information a besoin d'être explorée.

Compte tenu de ces observations, il semble important de préconiser que la visualisation pour la sécurité doit permettre aux experts de se concentrer le plus possible sur leurs objectifs (la sécurité) tout en les libérant des problèmes en dehors de leur milieu d'expertise (le design).

C'est dans ce but que nous avons conçu ELVis, un outil de visualisation de log pour la sécurité. ELVis permet aux experts en sécurité d'importer des fichiers de log ayant des formats multiples (par exemple, des fichiers de log apache standard et des fichiers syslog tels que les fichiers d'authentification) et de les explorer grace à des représentations pertinentes sélectionnées et générées automatiquement en fonction des données qui ont étés choisies.

Pour offrir ces fonctionnalités, ELVis identifie le format du log, transforme chaque ligne en champs et associe à chacun d'entre eux un type

de données. Ces types permettent d'une part d'enrichir les données avec des calculs descriptifs et d'autre part d'extrapoler des champs complémentaires. Ils permettent aussi faire d'associer automatique des visualisations appropriées aux champs sélectionnés.

⋈

L'analyse peut être considérée comme un processus de filtrage progressif que l'analyste adopte pour rechercher d'un élément spécifique d'information. Cependant, même si la détection de chaque action malveillante est fondamentale, il est aussi important de comprendre les relations entre les événements de sécurité pour pouvoir reconstruire le scénario global.

Une fois que l'analyste a trouvé un événement intéressant, il ou elle doit pouvoir découvrir les éléments connexes, même si ces événements se trouvent dans des fichiers de log différents, générés par différentes sources et donc exhibant des formats différents.

Quels sont, par exemple, les relations entre les attaques dans différents composants du système? Une fois qu'un serveur web est compromis, est-ce que les attaquants ont ensuite effectué d'autres actions malveillantes dans d'autres endroits du système? Dans ce cas là, quelles sont les conséquences? En réaction à ceci, nous affirmons que l'investigation en informatique est un processus itératif qui permet à l'analyste de facilement pouvoir utiliser les information stockées dans des fichiers de log, même si ceux-ci ne sont pas explicitement liés *a priori*.

C'est pour faciliter cette tâche que nous avons conçu CORGI, un outil web qui aider à explorer plusieurs fichiers log simultanément et qui permet à l'utilisateur de traverser plusieurs ensembles de données en suivant des points d'intérêt dans ces données.

Pour faire ceci de manière fiable, le système de typage utilisé pour associer des visualisations avec des sous-ensembles de données est étendu avec des types sémantiques qui permettent de découvrir des champs connexes entre les fichiers log. CORGI réutilise les capacités de visualisation de ELVis et les étend pour améliorer les possibilités d'exploration.

Le filtrage réactif des données est implémenté pour toutes les vues, et ces vues réagissent de façon synchronisée pour tout ensembles des données suivant l'interaction avec les points d'intérêt stockés. L'interaction utilisateur est une exploration guidée par des points d'intérêt et une interface conçue pour les cycles de recherche avec une approche de filtrages progressifs.

Les points d'intérêt peuvent non-seulement être utilisés pour filtrer et lier plusieurs logs, ils peuvent aussi donner des perspectives sur la progression et les résultats d'une analyse, fournissant ainsi l'essentiel pour partager des sessions et pour automatiquement générer des rapports.

ABSTRACT

Managing the vast quantities of data generated in the context of information system security becomes more difficult every day. Visualisation tools are a solution to help face this challenge. They represent large quantities of data in a synthetic and often aesthetic way to help understand and manipulate them.

In this document, we first present a classification of security visualisation tools according to each of their objectives. These can be one of three: monitoring (following events in real time to identify attacks as early as possible), analysis (the exploration and manipulation *a posteriori* of a an important quantity of data to discover important events) or reporting (representation *a posteriori* of known information in a clear and synthetic fashion to help communication and transmission).

We then present ELVis, a tool capable of representing security events from various sources coherently. ELVis automatically proposes appropriate representations in function of the type of information (time, IP address, port, data volume, etc.). In addition, ELVis can be extended to accept new sources of data.

Lastly, we present CORGI, an successor to ELVIS which allows the simultaneous manipulation of multiple sources of data to correlate them. With the help of CORGI, it is possible to filter security events from a datasource by multiple criteria, which facilitates following events on the currently analysed information systems.