BARING UNKNOWN ROGUE NETWORKS

La visualizzazione come strumento per analizzare
il comportamento dei network malevoli

Politecnico di Milano, Scuola del Design, Corso di Laurea Magistrale in Design della Comunicazione
Studente: Francesco Roveta 735719, Relatore: Paolo Ciuccarelli, Correlatore: Stefano Zanero
Anno Accademico 2010/2011

Alla mia famiglia
e agli amici.

Indice

Indice delle immagini

Figure

Tavole

Abstract

Nonostante esista una grande quantità di strumenti automatizzati utili all'individuazione di attività malevole, o sospettate come tali, all'interno della rete internet (come ad esempio attività di phishing, spam, o traffico botnet), l'analisi manuale condotta da parte degli esperti di sicurezza informatica è tutt'oggi ancora di fondamentale importanza nell'individuazione di elementi o eventi interessanti e nello studio delle dinamiche legate a queste tipologie di attività. Tuttavia questo compito si dimostra particolarmente difficile nel momento in cui si ha a che fare con attività lente, furtive, che lavorano nell'ombra e su larga scala. Per facilitare l'ispezione manuale di tutta questa serie di eventi, gli strumenti di visualizzazione sono un importante supporto in aiuto a esperti, analisti, e ricercatori del settore.

Il progetto presentato in questa tesi si propone come uno strumento interattivo di visualizzazione a supporto dell'analisi degli Autonomous System: quei gruppi di reti, controllati da una stessa organizzazione, che sono frequentemente caratterizzati dalla presenza al loro interno di attività sospettate di malevolenza. Nonostante il mercato attuale sia già ricco di prodotti e strumenti orientati in questa direzione, nella grande maggioranza dei casi si tratta di progetti sviluppati interamente da esperti del settore informatico, dei quali solamente una parte, e molto spesso in maniera non sufficiente, si preoccupa di sfruttare i mezzi forniti dalla visualizzazione per rendere al meglio i dati in proprio possesso.

Con la nostra proposta di tesi abbiamo voluto creare una collaborazione interdisciplinare tra due realtà differenti, la sicurezza informatica ed il design della comunicazione, con l'obiettivo di creare uno strumento il più possibile completo e funzionale, che potesse essere di aiuto nell'osservare, comprendere e tenere sotto controllo l'andamento delle attività malevole sulla rete.

(Eng) Despite the vast plethora of production-ready and automatic tools to detect suspicious or malicious activity on the Internet (e.g., phishing, spamming, botnet traffic), the manual analysis conducted by security experts are still fundamental to detect elements or interesting events and to study dynamics related to this activities. This task, however, is particularly hard when slow, stealthy and largescale activities are involved. In facilitating manual inspection of security related events, Visualization tools are a big support to help experts, analysts and researchers.

The project we present in our thesis is an interactive visualization tool to analize the Autonomous Systems: those groups of networks, controlled by the same organization, that are frequently characterized by activities suspected to be malicious. Despite a large number of analysis tools are currently available, the vast majority of these products are developed by domain experts (e.g., security researchers), strongly concentrated to the research results themselves and rarely oriented to visualization techniques, often not done in the better way to give a real added value to the results their analysis tool.

With our proposal of thesis we wanted to create a synergistic cooperation between two different disciplines: IT security and communication design. Our target was the creation of an as complete as possible tool, which really helps the experts to view, understand and keep under control malicious activity on the global net.

Introduzione

Questa tesi nasce dalla collaborazione tra due discipline differenti, ed in particolare dalla combinazione di esperienze nel campo dell'ingegneria e della sicurezza informatica, e nel campo del design della comunicazione e della data ed information visualization. Sviluppato in collaborazione con Luca Di Mario [1], studente in ingegneria informatica, il nostro progetto risponde ad una necessità espressa dal laboratorio di Valutazione delle Prestazioni e Affidabilità (VPlab) del Dipartimento di Elettronica ed Informazione (DEI) del Politecnico di Milano, che si occupa di valutazione di prestazioni, affidabilità e sicurezza dei sistemi informatici. Nello specifico il Politecnico ha recentemente preso parte ad un progetto di ricerca europeo chiamato WOMBAT, della durata complessiva di tre anni, il cui scopo era quello di creare, appoggiandosi sia ad università che a società internazionali operanti nel settore della sicurezza, una serie di strumenti e di prodotti utili ad analizzare, monitorare e contrastare il fenomeno delle minacce informatiche. All'interno di questa esperienza è stato creato uno strumento di raccolta dati chiamato FIRE, che si occupa di monitorare giornalmente la situazione mondiale relativa alle minacce informatiche, e raccogliere all'interno di un database tutte le informazioni sulle minacce rilevate. Attualmente però questa enorme quantità di informazioni è

[1] Per un approfondimento sulla parte di calcolo algoritmico fare riferimento alla tesi di laurea magistrale in ingegneria informatica di Luca Di Mario "BURN: Baring Unknown Rogue Networks - Studio e sviluppo di un'applicazione per l'analisi del comportamento delle reti malevole"

per lo più fruibile solamente sotto forma di enormi tabelle, e la ricerca manuale di fenomeni interessanti all'interno di questa enorme quantità di dati può risultare molto difficile e dispendiosa. Nasce da qui la necessità di riuscire a comunicare, tradurre al meglio ad un pubblico, sia di ricercatori ed esperti del settore, che di utenti meno esperti, questa grande quantità di informazioni, appoggiandosi ad uno strumento in grado di raccogliere grandi quantità di dati e informazioni in uno spazio ristretto, e capace di palesare in maniera semplice e immediata all'occhio umano eventi e fenomeni altrimenti difficilmente riconoscibili tra milioni di dati incolonnati in tabelle: la visualizzazione.

Partendo dai dati presenti nel database di FIRE, analizzandoli e riorganizzandoli, arricchendoli con nuovi dati ed algoritmi utili a individuare tendenze, fenomeni ed informazioni prima non esplicitate, il nostro progetto si è posto come obiettivo quello di creare uno strumento interattivo, basato su piattaforma web e liberamente fruibile, che attraverso la visualizzazione restituisse ad un utente finale questa grande quantità di informazioni in maniera chiara ed ordinata, e che fosse al tempo stesso sia utile per incuriosire ed avvicinare al problema un utente non esperto, sia strumento a supporto del lavoro di un esperto o ricercatore del settore.

"Whenever someone asks if anyone ever died in a cyber war, Magomed Yevloev springs to mind. On August 31, 2008, in the North Caucasus Republic of Ingushetia, Yevloev was arrested by Nazran police, ostensibly for questioning regarding his anti-Kremlin website Ingusheta.ru. As he was being transported to police headquarters, one of the officers in the car "accidentally" discharged his weapon into the head of Magomed Yevloev."

Jeffrey Carr, Inside Cyber Warfare
(O'Reilly, December 2009)

1 | Virus e altre minacce informatiche

"[…] the differentiation between Cyber Crime, Cyber Warfare and Cyber Terror can be a misleading one – in reality, Cyber Terror is often Cyber Warfare utilizing Cyber Crime."

Alexander Klimburg, Cyber-Attacken als Warnung
(DiePresse.com, 15 luglio 2009)

1.1 Nascita dei virus, una storia iniziata nel 1949

Nel suo concetto originario un virus altro non è che un insieme di istruzioni, un vero e proprio programma, specializzato per eseguire poche e semplici operazioni, in modo autonomo ed "invisibile" su di una macchina infettata.

La storia dei virus inizia nel 1949 quando John Von Neumann, matematico ed informatico ungherese, dimostrò in maniera del tutto teorica la possibilità di creare un programma per computer in grado di replicarsi autonomamente.
Basandosi sul principio dell'auto-replicazione nei primi anni 60 un gruppo di programmatori dei Bell Laboratories della AT&T ideò un gioco chiamato *Core Wars*, nel quale più programmi si dovevano sconfiggere sovrascrivendosi a vicenda. Era la prima evoluzione pratica dei concetti teorizzati da Neumann.
Circa vent'anni più tardi, nel 1984, Fred Cohen, informatico statunitense, utilizzò per la prima volta il termine "virus" nel suo scritto *Experiments with Computer Viruses*. In questo scritto Cohen indicava Leonard Adleman, matematico, informatico e biologo statunitense, come colui che aveva coniato tale termine. La definizione di virus era la seguente:

"Un virus informatico è un programma che ricorsivamente ed esplicitamente copia una versione possibilmente evoluta di sé stesso." [1]

Si tratta però soltanto di un primato in campo accademico, in quanto la parola

"Virus" legata al campo dell'informatica era già ampiamente diffusa nel linguaggio comune e nella letteratura, soprattutto di fantascienza.

Il primo vero virus per computer accreditato come tale viene attribuito a Rich Skrenta che nel 1982, all'età di quindici anni, creò un programma chiamato *Elk Cloner*. Questo programma girava sul DOS 3.3 della Apple e l'infezione era propagata tramite lo scambio di floppy disk.
Qualche anno più tardi, nel 1986, due fratelli pakistani proprietari di un negozio di computer misero in circolazione un virus chiamato Brain, creato per punire chi copiava in maniera illegale il loro software. Di lì a seguire, rispettivamente nel 1987, 1988, e 1989 fecero la loro comparsa i primi file infector, il primo worm ed i primi virus polimorfi, molto simili ai trojan dei giorni nostri.

Dopo una crescita e diffusione del fenomeno durata tutti gli anni '90, verso la fine del millennio, con l'avvento in massa di internet, la propagazione virale trova la sua strada definitiva: la rete. Internet e le email divengono il principale mezzo di diffusione per le più svariate tipologie di virus informatici.
Nel 2000 il famoso *I love you* diede il via al periodo degli script virus, diffondendosi attraverso la posta elettronica in milioni di computer, al punto da arrivare a coinvolgere una squadra speciale dell'FBI per l'arresto del suo creatore: un ragazzo delle Filippine che col suo programma aveva mandato in tilt i server di posta di tutto il mondo. Dal 2001 inoltre si è iniziato a registrare un incremento di worm che per diffondersi approfittano di falle di programmi o sistemi operativi senza bisogno dell'intervento dell'utente, portando in tal modo la velocità di diffusione di questi virus a livelli mai registrati prima. L'apice si ebbe nel 2004 quando il worm denominato *Slammer* in soli quindici minuti dal primo attacco infettò quasi la metà dei server che tenevano in piedi internet, mandando in tilt i bancomat della Bank of America, spegnendo il servizio di emergenza 911 a Seattle e provocando la cancellazione per continui inspiegabili errori nei servizi di biglietteria e check-in aeroportuali.

Attualmente riferirsi alle minacce informatiche mettendole tutte sotto la categoria

[1] F. Cohen, Computer viruses: theory and experiments. In Journal Computers and Security, Volume 6 Issue 1, Feb. 1987

"virus" può risultare fuorviante e spesso scorretto. Difatti oggigiorno i codici malevoli a cui si può attribuire il nome di virus sono pochi. Questo termine si riferisce a quella tipologia di programmi che essendo ancora legati per la loro diffusione ad un supporto di tipo fisico, dovevano essere piccoli, leggeri, performanti ed il più silenziosi possibili. Virus era quindi un piccolo codice malevolo, difficile da individuare e, soprattutto, per sopravvivere doveva riuscire a diventare parte integrante di un altro programma, infettandolo. Al giorno d'oggi le informazioni viaggiano da un capo all'altro del mondo ormai senza vincoli di tipo fisico, e se questi vincoli ci sono l'utente medio non se ne preoccupa più. La materia si è fatta informazione, e l'informazione non ha un luogo fisico di riferimento. Inoltre le macchine sono sempre più performanti e la banda larga è ormai alla portata di tutti.
Questo insieme di fenomeni può portarci a sostituire il vecchio concetto di virus col più moderno concetto di worm: programmi scritti con linguaggi di livello sempre più elevato che lavorano sfruttando le vulnerabilità dei sistemi operativi, primo tra tutti windows. La possibilità di utilizzare linguaggi di programmazione di diverso tipo porta ad una diversificazione ed ad un aumento significativi delle tipologie di programmi malevoli in circolazione. Inoltre questi ultimi vivono in maniera a se stante, non hanno più bisogno di legarsi ad un altro programma per girare, non sono soltanto parti di codice malevolo, sono veri e propri applicativi a tutti gli effetti. Invece che infettare i file, operazione che richiedeva più tempo e complessità, questa nuova tipologia di programmi si replica passando di macchina in macchina senza nemmeno preoccuparsi troppo di nascondere la propria attività.

Ci troviamo quindi di fronte a nuove tipologie di codici malevoli, potenzialmente sempre più potenti e complessi, che sfruttano la rete e le sue debolezze per moltiplicarsi a velocità mai raggiunte prima, per raggiungere un maggior numero di potenziali bersagli.

1.2 Un'evoluzione naturale, la nascita di una fiorente economia criminale

Nel corso degli anni, l'evoluzione naturale di tutto questo è stata la nascita di una sempre più fiorente economia criminale attorno ad attività illecite operanti

tramite la rete, da frodi legate al furto di identità, all'intrusione in conti correnti e carte di credito online, spam, phishing e così via. Col tempo, come spesso la storia ha insegnato quando ci sono due forze opposte in gioco, il campo della sicurezza informatica è cresciuto di pari passo con la crescita di questi fenomeni, vivendo in dipendenza del suo avversario.

"The motivation for security research is ever to stymie the goals of some hypothetical miscreant determined to violate one of our security policies. Typically, we abstract away their motivations and consider the adversary solely in terms of their capabilities. There is good reason for this since the threat model for any security mechanism is generally driven entirely by the adversary's abilities. Moreover, reasoning about any individual's state of mind, let alone predicting their behavior, is inherently prone to error. That said, the nature of Internet-based threats has changed over the last decade in ways that make it compelling to attempt a better understanding of today's adversaries and the mechanisms by which they are driven." [2]

Il principale obiettivo di un ricercatore che opera nel campo della sicurezza informatica è quindi da sempre quello di ostacolare le attività di ipotetici possibili malfattori. Tipicamente, quando si ha a che fare con questo tipo di attività, si tende a prescindere quelle che possono essere le motivazioni che stanno dietro a determinate azioni, per concentrarsi esclusivamente sul come queste azioni vengano messe in atto, e di conseguenza sulle abilità e capacità tecniche di chi le sta perpetrando. Questa linea di pensiero caratterizza gran parte dei sistemi di sicurezza attualmente esistenti, con valide ragioni a supporto. Recentemente però la natura di queste minacce sta cambiando in maniera significativa, rendendo necessario un cambio di rotta anche dal punto di vista dell'approccio alla sicurezza, iniziando a rendere interessante il tentativo di comprendere più a fondo la natura di chi commette un'azione con finalità malevole e quali motivazioni lo portano a fare ciò.

Tra i vari cambiamenti che hanno interessato il fenomeno delle minacce informatiche, quello più importante è stato il suo graduale spostamento da un'economia basata sulla reputazione, dove l'appagamento personale veniva dal riconoscimento pubblico delle proprie capacità, ad un'economia basata

[2] J. Franklin, V. Paxson, A. Perrig, and S. Savage. An Inquiry into the Nature and Causes of the Wealth of Internet Miscreants. In ACM Conference on Computer and Communication Security (CCS), 2007

sul denaro, attraverso attività ad esempio di spam (invio di grandi quantità di messaggi spesso a scopo pubblicitario), phishing (furto di identità attraverso finti messaggi o finti siti web) o estorsione tramite attacchi DDoS (Distributed Denial of Service). Inoltre, anche attività legali legate alla ricerca sulle vulnerabilità dei sistemi, hanno in alcuni casi cambiato fazione, entrando a far parte di questa sempre più fiorente economia, ed oggigiorno esiste un vero e proprio mercato attorno alle informazioni sulle vulnerabilità, regolarmente acquistate e vendute sia da aziende pubbliche che da organizzazioni private, e segrete [3].

Contrariamente a quanto accadeva diversi anni fa, oggigiorno è quindi utile andare ad indagare sui comportamenti e sulle motivazioni che spingono a questo tipo di attività, una vera e propria economia criminale basata sugli attacchi di tipo informatico. Inoltre si può parlare di criminalità informatica in quanto il fenomeno ha superato i confini del singolo gruppo che agiva in maniera chiusa ed individuale, arrivando addirittura all'esistenza di un traffico online di beni e servizi a supporto di queste attività. Così, mentre rimane difficile tentare di analizzare le cause e le motivazioni che stanno dietro un singolo attacco informatico, come anche risalire all'autore dello stesso, si può iniziare ad analizzare gli andamenti generali di questo mercato e le forze che in esso sono in gioco.

1.3 La mancanza di una legislazione unitaria, Cyber Crime e Cyber Warfare

Per fare un ulteriore passo avanti nell'analisi del problema delle minacce informatiche si rende necessario un approfondimento ed una breve introduzione su quello che viene chiamato cyber crime e, ad un livello più elevato, cyber warfare.

Quando ci chiediamo quale portata possa avere una minaccia informatica nella vita reale di tutti i giorni, nei rapporti tra le persone, o tra nazioni, dobbiamo tenere conto del fatto che ormai i computer, l'informatica, internet, sono diventati parte integrante della nostra esistenza, hanno cambiato il nostro modo di vivere

[3] Charlie Miller. The legitimate vulnerability market: Inside the secretive world of 0-day exploit sales. In Sixth Workshop on the Economics of Information Security, May 2007

e vedere il mondo, e soprattutto il nostro modo di comunicare. Inoltre sempre più dati e informazioni, e per informazioni si intendono anche le informazioni riservate appartenenti a stati, governi, e così via, si riversano giornalmente nella rete, in database digitali e supporti informatici, e, come ci ha insegnato la recente vicenda di WikiLeaks, possono diventare potenti armi.
A tutto questo va aggiunto il fatto che tutt'ora non esiste una regolamentazione unitaria a livello mondiale che concordi sulla linea da tenere riguardo alla problematica degli attacchi informatici, o sul che cosa sia considerabile attacco informatico. Leggi e regolamentazioni cambiano stato per stato, e la natura stessa della rete permette ai criminali informatici di gestire server sparsi per tutto il globo, che fisicamente si trovano in luoghi con legislazioni spesso completamente diverse tra loro, aggiungendo alla difficoltà fisica di individuare queste attività la difficoltà burocratica del come agire per fermare, o perlomeno rallentare, questo fenomeno.

Sempre per dare un'idea della portata che può raggiungere, nel Virtual Criminology Report stilato da McAfee nel 2008 si può leggere che "esistono più di 120 nazioni nel mondo che utilizzano internet per attività di spionaggio in campo politico, militare ed economico". Altre informazioni interessanti sull'argomento si possono trovare nel libro *Inside Cyber Warfare, mapping the cyber underworld* di Jeffrey Carr [4], libro in cui l'autore riporta alcuni interessanti esempi sulla situazione della cyber warfare nel ventesimo e ventunesimo secolo. Riporto qui di seguito alcuni esempi tratti da questo libro.

Cina

La nascita di una comunità di hacker ufficialmente appartenenti alla Repubblica Popolare Cinese (PRC) può essere stimata intorno al 1998. Circa un anno più tardi, il 7 maggio 1999, un jet della NATO bombardò per errore l'ambasciata cinese di Belgrado, in Yugoslavia. Meno di 12 ore più tardi iniziarono una serie di attacchi contro centinaia di siti governativi degli USA, rivendicati da quella che si fece chiamare la Chinese Red Hacker Alliance.
Due anni più tardi, nel 2001, un jet militare cinese ebbe una collisione con un aircraft militare statunitense sui mari a sud della Cina. Questa volta oltre 80.000 hackers vennero ingaggiati in un progetto di "autodifesa" contro quello che ven-

[4] Jeffrey Carr. Inside Cyber Warfare: Mapping the Cyber Underworld. O'Reilly, December 2009

ne definito un atto di aggressione da parte degli USA. Il *New York Times* definì questi avvenimenti una "World Wide Web War I".
Da allora gran parte dell'attività di "cyber espionage" condotta dalla Repubblica Popolare Cinese si è concentrata sul tentare di colmare il divario tecnologico in campo militare tra loro e gli USA.

Israele

Nel dicembre 2008 Israele lanciò l'operazione Cast Lead contro la Palestina. In breve tempo una vera e propria guerra informatica esplose tra Israele e gli hackers arabi. L'aspetto che rende importante questo avvenimento è il fatto che gli attacchi vennero portati da entrambi gli stati attraverso hackers ufficialmente appartenenti ad essi, invece che da non-state hackers, pratica più comunemente usata, per evitare di far risalire direttamente gli attacchi ad uno stato di provenienza. Membri della Istraeli Defence Force si introdussero all'interno del canale televisivo palestinese Al-Aqsa, stazione televisiva ufficiale di Hamas, e vi inserirono un cartone animato che mostrava la morte dei leader di Hamas con in sovraimpressione la scritta in arabo *"il tempo sta per scadere"*.

Russia

La seconda guerra russo-cecena (1997-2001). Durante questo conflitto, in cui le forze militari russe invasero alcune regioni cecene con l'obiettivo di reinstallarvi un regime pro-moscovita, da entrambi i lati vennero utilizzati attacchi informatici mirati a operazioni di tipo informativo per esercitare un controllo sull'opinione pubblica. Anche dopo la fine ufficiale della guerra, la Russian Federal Security Service (FSB) venne considerata responsabile per aver messo fuori gioco due importanti siti ceceni nello stesso momento in cui la Russian Spetsnaz Troop ingaggiava i combattimenti con il gruppo di terroristi ceceni che il 26 ottobre 2002 tenne in ostaggio diversi civili russi in un teatro a Mosca.
The Estonian cyber war (2007). Anche se non sono mai stati ufficialmente trovati collegamenti tra il governo russo e gli attacchi informatici lanciati contro i siti governativi dell'Estonia nell'aprile del 2007, un importante giovane leader della Russian Nashi, Konstantin Goloskokov, ammise il suo coinvolgimento nei fatti insieme ad alcuni altri colleghi. La causa dell'incidente fu la decisione da parte dell'Estonia di spostare la statua *The Bronze Soldier of Tallinn*, dedicata agli

ex soldati dell'Unione Sovietica morti in quella battaglia. In tutta risposta un attacco DDoS (Distributed Denial of Service) portò al crollo di diversi siti estoni appartenenti a banche, parlamento e ministeri.

The Russia-Georgia War (2008). Questo è il primo esempio di un attacco informatico direttamente coincidente con un'invasione fisica via terra, mare ed aria condotta da uno stato contro un altro. La Russia invase la Georgia in risposta agli attacchi separatisti condotti da quest'ultima nel sud dell'Ossezia. L'attacco informatico, perfettamente coordinato con quello fisico, prese di mira importanti siti governativi georgiani e tutta una serie di altri siti strategicamente importanti, come quelli appartenenti alle ambasciate statunitensi e britanniche.

Iran

Le elezioni presidenziali iraniane del 2009 portarono ad una protesta di massa abbastanza importante contro i brogli elettorali avvenuti, questa protesta fu in gran parte alimentata dalla possibilità della popolazione di condividere le informazioni attraverso la rete e soprattutto tramite social networks quali Twitter e Facebook. Il governo iraniano rispose istituendo una linea di azione dura contro i protestanti, limitando l'accesso a Internet all'interno del paese e tentando di bloccare tutti quei social media che in breve tempo divennero l'unico mezzo di comunicazione e informazione sia tra le persone coinvolte nella protesta che verso gli altri stati. Alcuni membri dei movimenti di protesta iniziarono inoltre a lanciare attacchi DDoS contro i siti governativi iraniani ed ad utilizzare Twitter come mezzo per reclutare altri "cyber warriors" che si unissero alla loro causa, linkando nei loro messaggi software DDoS automatizzati per permettere a tutti di partecipare attivamente alla loro azione.

Korea del nord

Il 4 luglio 2009 qualche dozzina di siti statunitensi, tra cui quello della Casa Bianca e altri siti governativi, finirono nel mirino di un attacco DDoS. Qualche giorno più tardi la lista dei siti sotto attacco comprendeva anche alcuni siti sudcoreani, sia governativi che civili. La prima sospettata dell'attacco fu ovviamente la corea del nord, anche se non se ne ebbe mai la conferma evidente. Nonostante questo la Corea del Sud spinse per ottenere da parte dell'esercito statunitense un attacco informatico ai danni della Corea del Nord come avvertimento.

Come abbiamo visto si tratta quindi di un fenomeno dalle molteplici sfaccettature, e che si è evoluto e sviluppato a diversi livelli di importanza, dalla frode al singolo individuo fino ad attacchi coordinati contro intere nazioni. Attualmente non vi è ancora una linea comune a livello globale sul come relazionarsi con questo tipo di attività. Difatti oltre alle problematiche legate all'individuazione dei responsabili di un attacco bisogna prendere in considerazione anche tutte le problematiche relative al luogo fisico da dove l'attacco è partito, relazioni internazionali, leggi diverse da uno stato all'altro, difficoltà di condurre ricerche all'interno di un altro stato, e così via. Si tratta di attacchi che avvengono attraverso il mezzo informatico, che tramite la rete possono colpire in qualsiasi parte del globo, ma che al giorno d'oggi fanno ancora capo ad un sistema legislativo strettamente legato alle regole vincolate dai confini fisici. Inoltre a tutto questo si aggiunge la difficoltà di identificare un attacco nel momento in cui questo sta avvenendo: solitamente le registrazioni di un'attività malevola avvengono quando l'attacco è già stato sufficientemente avviato o concluso.

Per tutte queste motivazioni, e non solo, la tendenza comune in materia di sicurezza informatica è quella di agire sulla difensiva, tentando di individuare nella maniera più precisa e nel minor tempo possibile eventuali attacchi, con una politica della limitazione dei danni, e, una volta individuati, mettendo alla berlina i criminali informatici, nella speranza che le autorità di competenza prendano adeguati provvedimenti a riguardo.

2 | Il bisogno di visualizzare

"Why should we be interested in visualization? Because the human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain from a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers."

Colin Ware, author of Information Visualization: Perception for Design

La rappresentazione visiva dei dati ci permette di comunicare a chi osserva una grande quantità di informazioni. Troppo spesso le informazioni si trovano soltanto sotto forma di testo ed è più difficile cogliere nell'immediato l'essenza di qualcosa se questa è espressa solamente tramite le parole. Difatti è più difficile per il nostro cervello processare un testo, mentre figure e immagini necessitano di un tempo più breve per essere comprese ed assimilate, e possono contenere grandi quantità di informazioni. Inoltre le immagini possono utilizzare forme, colori, dimensioni, posizione, movimento e tutta un'altra serie di attributi per codificare l'informazione, contribuendo a fornire a chi osserva una grandissima quantità di dati in spazi relativamente ristretti. Molte discipline si stanno scontrando con una sempre crescente quantità di dati che necessitano di essere analizzati, processati e soprattutto comunicati. Siamo nel mezzo di un'era di esplosione dell'informazione e attualmente gran parte di quest'ultima è immagazzinata o comunicata esclusivamente sotto forma di testo: database, documenti, quotidiani, e così via.

In questo panorama inizia a sentirsi la necessità di trovare nuove strade per rapportarci a questa crescente quantità di informazione che ci circonda. Abbiamo bisogno di mezzi per trasferire l'informazione da scritta a visiva, per aiutarci nei processi di comprensione dei dati e nei processi di analisi degli stessi. Avere la possibilità di visualizzare grandi quantità di dati è una risorsa fondamentale per trovare informazioni. E l'interazione con essi è un altro elemento chiave. Spesso

la rappresentazione visiva, rispetto alla rappresentazione testuale, aiuta a riconoscere relazioni, andamenti, casi particolari altrimenti difficilmente individuabili. Questa necessità di visualizzazione, sentita in molti ambiti e discipline differenti, ha iniziato ad interessare sensibilmente anche a chi si occupa giornalmente di sicurezza informatica, e la grande quantità di dati che ogni giorno vengono sfornati da report e database di sicurezza fornisce delle ottime basi su cui lavorare.

2.1 Come sfruttare al meglio il sistema visivo umano, i vantaggi della visualizzazione

Come abbiamo già detto la visualizzazione può offrire notevoli vantaggi rispetto all'analisi testuale dei dati. Questi vantaggi nascono dalla capacità di processo delle immagini della mente umana. Le persone possono visionare, riconoscere e ricordarsi molto più rapidamente le immagini rispetto ad un testo. Inoltre la mente umana è particolarmente abile nel riconoscere ed individuare cambiamenti nelle forme, colori, movimenti, dimensioni, e così via. In accordo con quanto scritto da Raffael Marty nel suo libro Applied Security Visualization [1], la visualizzazione può diventare un potente mezzo per l'analisi per svariati motivi:

Rispondere alle domande: tramite la visualizzazione è possibile creare un'immagine per ogni domanda che si possa avere circa un dato set di dati. Al posto che tentare di districarsi attraverso grandi quantità di dati testuali, cercando di ricordarsi tutte le relazioni esistenti tra essi, si può utilizzare la visualizzazione per convergere tutti i dati in un'unica immagine.

Creare nuove domande: aspetto molto interessante che riguarda la visualizzazione è la sua capacità di generare in chi la guarda nuove domande. La mente umana quando messa di fronte ad un'immagine automaticamente inizia ad individuarne andamenti, pattern ed incongruenze. Spesso tutto ciò risulta difficile da individuare all'interno di un file di testo.

[1] Raffael Marty. Applied Security Visualization. Addison-Wesley, August 2008

Esaminare e scoprire: la visualizzazione dei dati non è mai univoca, diverse visualizzazioni dello stesso set di dati possono metterne in evidenza aspetti diversi. Questi erano sicuramente presenti anche in precedenza, ma senza un ordine visivo risultano decisamente più difficili da individuare. La visualizzazione può essere un ottimo strumento per la scoperta, ed unita all'interazione accresce la sua capacità indagativa.

Sostenere decisioni importanti: la visualizzazione aiuta nell'analisi di grandi quantità di dati in un tempo molto ristretto. Questo può portare ad avere sott'occhio un quadro generale della situazione, e a distinguerne gli aspetti più interessanti. La situational awareness, ovvero *"the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future, […]"* (Endsley, 1995), è uno strumento essenziale a supporto dei processi decisionali.

Comunicare l'informazione: la rappresentazione visiva dei dati, dal punto di vista comunicativo, può essere molto più efficiente rispetto ad una rappresentazione testuale degli stessi. Il tempo necessario per comprendere un'immagine può essere molto più breve rispetto a quello richiesto per la comprensione di un testo, senza per questo andare a perdere di informazione o significato.

Ispirare: le immagini creano ispirazione. Analizzare rappresentazioni visive porta a cercare di fare nuove considerazioni su quanto stiamo vedendo, a pensare se non ci siano modi migliori per visualizzarle ed analizzarle, per mettere in evidenza nuove relazioni, o semplicemente per cambiarne il punto di vista.

Inoltre non bisogna dimenticare che nel panorama di questa disciplina si sente ancora molto la necessità di affiancare ai meccanismi di detection automatizzati, l'analisi manuale dei dati da parte di esperti del settore, essenziale per indagare al meglio il fenomeno. Questo bisogno diviene particolarmente impegnativo nel momento in cui l'analista si deve scontrare con le strategie tipiche dei moderni criminali informatici, che prevedono attività che si prolungano a lungo nel tempo, che comprendono spesso azioni su larga scala, saltando da un punto all'altro del globo, e che ovviamente perpetrano tentando con tutti i mezzi possibili di

nascondere la propria identità. Spesso l'analista si trova quindi di fronte a vere e proprie montagne di dati, tra i quali districarsi può rivelarsi un'impresa particolarmente ostica. Per facilitare questo tipo di lavoro, una corretta visualizzazione dei dati può portare un sensibile aiuto.

2.2 Cos'è e di cosa si occupa la Security Visualization

Quando parliamo di Security Visualization parliamo di un ambito di interesse davvero recente. Nel campo della sicurezza informatica ci si trova spesso ad avere a che fare con ingenti quantità di dati che necessitano di essere analizzati per risolvere problemi di sicurezza, prevenirli, o semplicemente monitorare lo stato del proprio network. A causa della sempre crescente quantità di dati prodotti e collezionati gli strumenti più comuni, come i firewall o i sistemi di monitoraggio delle intrusioni, non sono più sufficienti. I classici report non bastano più per studiare e tentare di prevenire il fenomeno, ne tantomeno le aggiunte grafiche che recentemente hanno iniziato a comparire nelle dashboard di un gran numero di questi programmi: la maggior parte di queste visualizzazioni si limita a tradurre in grafici esigue quantità di dati, in maniera molto elementare, e senza preoccuparsi di fornire un approccio interattivo che aiuti l'utente nei processi di analisi.
I prodotti per la sicurezza tutt'oggi non sono ancora progettati con un occhio sufficientemente critico alla visualizzazione, anche se ultimamente si sta iniziando a vedere un certo interesse all'argomento ed una evoluzione in questa direzione. Le aziende che si occupano di sicurezza iniziano a realizzare che la visualizzazione può trasformarsi in importante vantaggio competitivo e un grande aiuto per l'utente.

Sempre rifacendomi agli scritti di Raffael Marty, data la grande quantità di dati che necessita di essere analizzata nell'ambito della sicurezza informatica, la visualizzazione degli stessi sembra essere il giusto approccio per i seguenti motivi:
1. La sempre maggior quantità di dati collezionati nel campo dell'Information Technology necessita di nuovi metodi e strumenti per la loro analisi;
2. L'analisi di eventi e log sta diventando uno degli strumenti principali in mano agli analisti per investigare e comprendere lo stato della sicurezza dei propri network, host, applicazioni, etc. Tutto questo lavoro ha a che fare con una

grande quantità di dati che necessita di essere analizzata;
3. Il panorama della criminalità si sta trasformando. Si può ormai parlare tranquillamente di criminalità informatica, attacchi informatici, frode e sottrazione di informazioni sul web, etc. Tutta questa attività genera una gran quantità di dati che va raccolta e analizzata.

2.2.1 Raffael Marty

Esperto nel campo della sicurezza informatica, Raffael Marty è il fondatore di PixlCloud [2] e co-fondatore di Loggly [3]. Il suo campo di ricerca spazia in tutto quello che può essere correlato all'IT data visualization. Secviz.org [4], di cui è fondatore, è uno dei più importanti portali attualmente esistenti dedicati al campo della security visualization. Ulteriore contributo alla materia viene anche dal suo libro *Applied Security Visualization* [1]. Il libro si presenta come una guida all'analisi visuale dei dati legati all'ambito della sicurezza informatica. Al posto di concentrarsi su dati prettamente testuali, raccolti in tabelle, la visualizzazione degli stessi offre migliori possibilità di ricerca e analisi. Una visualizzazione grafica aiuta immediatamente a individuare comportamenti anomali, attività sospette e incongruenze negli andamenti, o ad inviduare trend e relazioni all'interno di un dataset. Per utilizzare le parole dell'autore:

"Visualization of data - the process of converting security data into a picture - is the single most effective tool to address these tasks".

2.2.2 secviz.org

Secviz.org è un portale nato nel 2006, fondato da Raffael Marty e pensato per tutti coloro che si occupano di log analisys, log mining e di visualizzazione legata al mondo della sicurezza informatica. Il portale si pone come luogo dove discutere, scambiarsi opinioni ed informazioni, dividere conoscenze, tecniche, metodi ed esempi di grafici e visualizzazioni sull'argomento.

Nella sezione dedicata ai grafici gli utenti hanno la possibilità di inserire immagini e dettagli relativi alle proprie ricerche nel campo della visualizzazione, o link

[2] http://pixlcloud.com/
[3] http://www.loggly.com/
[4] http://secviz.org/

a tool o software interessanti trovati in rete.

Tutti i più interessanti tool gratuiti presenti sono inoltre raccolti e disponibili all'interno di un cd, DAVIX live CD, che si trova in allegato al libro Applied Security Visualization, o si può scaricare gratuitamente dal sito.

2.2.3 vizsec.org

Portale anch'esso nato nel 2006, gestito da Secure Decisions [5], divisione di Applied Visions Inc. [6], ospita una comunità di ricerca e sviluppo che lavora sull'applicazione delle tecniche di Information Visualization ai problemi della sicurezza informatica. Il sito oltre a contenere un forum di discussione, raccoglie, suddivisi per tipologie, un gran numero di toolkit, software ed applicazioni per la visualizzazione.

Inoltre dal 2004, due anni prima della nascita del portale, vengono organizzati dei workshop annuali sulla visualizzazione in ambito di sicurezza informatica. Durante questi eventi ricercatori, studenti, esperti del settore, società, presentano e condividono soluzioni alle moderne sfide della cyber security utilizzando metodi di visualizzazione. Dall'edizione del 2011 VizSec si è trasformata da workshop a symposium, con l'obiettivo di diventare nei prossimi anni vera e propria conferenza, a sottolineare l'importanza che questi incontri stanno guadagnando nell'ambito della Cyber Security.

[5] http://www.securedecisions.com/
[6] http://www.avi.com/

3 | Visualizzare le minacce informatiche: l'attuale stato dell'arte

"Manual analysis of security-related events is still a necessity to investigate non-trivial cyber-attacks. This task is particularly hard when the events involve slow, stealthy and large-scale activities typical of the modern cyber-criminals' strategy. In this regard, visualization tools can effectively help analysts in their investigations."

In "BURN: Baring Unknown Rogue Networks."
(In Proc. of the 8th Intl. Symposium on Visualization for Cyber Security, VizSec '11)

Per molti anni i criminali informatici hanno potuto condurre le loro attività illecite nascondendosi dietro Internet Service Providers (ISPs), coloro che gestiscono le connessioni, che non si preoccupavano di tenere sotto controllo il fenomeno, anzi in molti casi erano più interessati a mantenerlo attivo, spesso anche per proprio tornaconto. Una delle caratteristiche principali che contraddistingue questi ISPs è la longevità significativa che alcune attività malevole hanno all'interno del loro network, indice di un apparente mancanza di controllo, o connivenza, su ciò che avviene sui server connessi tramite la loro rete. Una nota interessante è che nonostante Internet, per sua stessa natura, garantisca un certo buon grado di anonimato, questi ISPs sono spesso abbastanza attenti alla loro immagine pubblica. Una volta esposti, nella maggior parte dei casi si nota una cessazione dell'attività malevola, o un de-peering (sconnessione) dei server, appartenenti alla rete, individuati come malevoli. Nonostante esistano un gran numero di sistemi automatizzati per la detection di questo genere di attività, l'analisi manuale da parte di esperti di sicurezza informatica è tutt'oggi ancora una necessità profonda, per una migliore comprensione del fenomeno. Nel corso degli anni sono stati sviluppati molteplici sistemi di visualizzazione legati all'argomento, spesso costruiti più da esperti del settore che da esperti nel campo della comunicazione. In questo capitolo ho fatto una raccolta dei principali sistemi attualmente esistenti.

3.1 La ricerca e le pubblicazioni scientifiche

Il campo della ricerca scientifica è forse quello più spinto verso lo studio e l'individuazione di nuovi mezzi e metodi per individuare, rallentare, contrastare il fenomeno. Esiste una vasta bibliografia sull'argomento, supportata e continuamente arricchita dalla costante collaborazione tra diverse università e laboratori di ricerca e dall'esistenza di numerose conferenze sull'argomento. Qui di seguito riporto tre esempi presi tra i più recenti sistemi sviluppati in questo campo.

3.1.1 NICTER: An incident analysis system

NICTER [1], il *Network Incident analysis Center for Tactical Emergency Response*, è un sistema progettato per monitorare ed analizzare in maniera del tutto automatica grandi quantità di dati ed attività presenti in rete, con l'obiettivo di correlare i fenomeni osservati con le loro possibili cause, per contribuire nella ricerca contro i problemi di sicurezza informatica causati dai malware. Questo sistema, avvalendosi di varie tecniche di analisi e data mining, mira ad offrire agli operatori del settore uno strumento in grado di:

1. Individuare anomalie o cambiamenti nel traffico dati monitorato, più velocemente di quello che potrebbe fare manualmente un operatore umano;
2. Classificare i server osservati come malevoli in base ai loro comportamenti;
3. Offrire una predizione su possibili cambiamenti importanti nel traffico dati, diverse ore prima che questi si verifichino.

Per fare questo NICTER si avvale di quattro sistemi integrati di detection (the Macro analysis System, the Micro analysis System, the Network and malware enchaining System, the Incident Handling System), focalizzati su una precisa tipologia di attività malevola (i malware), che permettono un'analisi della rete sia dal punto di vista macro che nel dettaglio. L'analisi macroscopica si basa sul network monitoring e si concentra sul tentare di comprendere i trend attuali delle attività malevole presenti all'interno dei diversi network. L'analisi nel dettaglio si focalizza sui singoli malware, andandone a studiare la struttura ed

[1] D. Inoue, K. Yoshioka, M. Eto, M. Yamagata, E. Nishino, J. Takeuchi, K. Ohkouchi, and K. Nakao. An incident analysis system nicter and its analysis engines based on data mining techniques. In Proc. of the 15th intl. conference on Advances in neuro-information processing - Volume Part I, ICONIP'08, pages 579-586, Berlin, Heidelberg, 2009. Springer-Verlag.

il funzionamento, per ottenere maggiori informazioni sulle loro caratteristiche.

Macro analysis System (MacS) Utilizza una serie di sensori per monitorare in tempo reale le darknet. Una darknet è un set di indirizzi IP globalmente riconosciuti come non utilizzati. Nel momento in cui vi è del traffico in arrivo da questi IP si può supporre che, poiché, l'essere inseriti in una darknet, indica che non vi sono host autorizzati ad utilizzare questi indirizzi, si tratti o di traffico proveniente da una qualche tipologia di attività malevola, o di un errore. Tutto il traffico proveniente da questi IP viene quindi monitorato e analizzato.

Micro analysis System (MicS) Tramite l'utilizzo di honeypots (settati su diversi livelli di sensibilità), cattura i malware presenti in rete e li gira automaticamente ad un apposito analizzatore, per osservarne il codice ed estrarne le caratteristiche ritenute più interessanti.

Network and malware enchaining System (NemeSys) Serve per concatenare il fenomeno (il singolo evento), e la sua possibile causa (un malware). Una volta che tramite il MacS viene individuato un host malevolo, NemeSys inizia a ricercare una possibile lista di malware con caratteristiche congruenti con il comportamento della rete che ospita l'host malevolo.

Incident Handling System (IHS) Serve per aiutare l'operatore ad osservare i risultati ottenuti tramite le analisi precedentemente descritte e stilarne un report. Tutti questi sistemi prevedono un resa dei risultati anche visiva attraverso singoli grafici specifici per ogni tipologia di dati, per aiutare l'operatore nell'analisi e nel report degli stessi.

3.1.2 EMBER: A Global Perspective on Extreme Malicious Behavior

EMBER [2], Extreme Malicious Behaviour viewER, è un sistema per l'analisi, la geolocalizzazione e la visualizzazione delle attività malevole registrate in rete. Spesso, per rendere visivamente la grande quantità di dati che vengono trovati in queste tipologie di ricerche, vengono utilizzati svariati tipi di visualizzazioni e grafici, la qual cosa da un lato può sicuramente fornire un significativo aiuto ad operatori, ricercatori ed esperti del settore, dall'altro, se non gestita correttamente, può trarre in inganno eventuali utenti meno esperti, ma non solo.
Una tipologia di visualizzazione che molto spesso è soggetta a questo tipo di errori è quella geografica, usata in questo campo per localizzare sul territorio la posizione di reti e server malevoli.

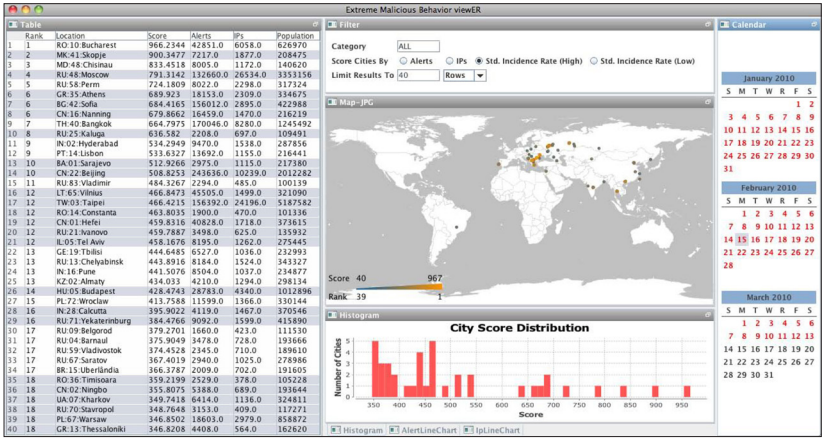
Il progetto EMBER concentra la sua attenzione proprio su quest'ultima, sottolineando come il segnalare su di una mappa, attraverso svariate tipologie di puntatori, la posizione dei diversi server, o colorare le regioni semplicemente in base all'intensità dell'attività in esse registrata, può portare alla creazione di visualizzazioni disordinate, che enfatizzano quelle aree dove la maggior densità di popolazione, e quindi la maggior presenza di connettività alla rete, portano ad una naturale presenza di un maggior numero di attività malevole.
Per ovviare a questo problema, e andare a dare il giusto peso ai dati che i sistemi di detection ci forniscono, vi è la necessità di andare a normalizzare i dati inerenti le attività registrate coi dati relativi alla quantità di popolazione e computer presenti nell'area.

Per fare questo EMBER propone l'utilizzo di un'unità di misura chiamata *Standardized Incidence Rate* (SIR), che misura il numero di server contenenti attività malevola per ogni 100.000 server disponibili in quella determinata area geografica, basandosi sulla localizzazione geografica dei diversi indirizzi IP, l'attuale densità di popolazione presente e la percentuale di utilizzo dei computer nell'area. Questo consente di individuare quelle aree che sono caratterizzate sia da un livello molto alto di attività malevola, le aree a rischio, sia

[2] T. Yu, R. Lippmann, J. Riordan, and S. Boyer. Ember: a global perspective on extreme malicious behavior. In Proc. of the Seventh Intl. Symposium on Visualization for Cyber Security, VizSec '10, pages 1-12, New York, NY, USA, 2010. ACM.

da uno molto basso, le aree presumibilmente ben protette, ed osservare quali zone sono maggiormente nel mirino di alcune specifiche tipologie di attacchi e quali invece vengono preferibilmente evitate.

Oltre a questa attenzione al trattamento dei dati EMBER tenta anche di fornire una resa grafica dei risultati ottenuti attraverso una apposita dashboard. La schermata centrale di quest'ultima contiene una mappa geografica dove le attività malevole sono raggruppate per città, e rappresentate su di essa come piccoli cerchi di dimensione e colore variabili, ad indicare la quantità di attività malevola, normalizzata, riscontrata in quella città. Sulla destra è presente un calendario, che fornisce la possibilità di selezionare quale giorno andare ad analizzare. Un pannello posizionato sopra la mappa consente di andare a personalizzare alcuni parametri per l'analisi, mentre un altro pannello posizionato sotto la mappa propone un istogramma riassuntivo della situazione globale. Infine sulla sinistra una tabella fornisce alcuni ulteriori dati sulle singole città.



▲ fig 01 | Interfaccia di EMBER
Immagine tratta dal paper Ember: a global perspective on extreme malicious behavior.

3.1.3 FIRE: FInding Rogue nEtworks

FIRE [3], acronimo di FInding Rogue nEtworks, è un sistema che mira ad identificare ed esporre tutte quelle organizzazioni ed ISPs (Internet Service Providers) che dimostrano un persistente atteggiamento malevolo sulla rete. L'obiettivo è quello di identificare tutti quei network che sono costantemente implicati in attività illecite. A questo scopo FIRE giornalmente monitora l'attività della rete e ne raccoglie i dati, inserendoli in un database, per poi renderli pubblici attraverso il sito maliciousnetworks.org.
Come abbiamo già detto l'analisi delle attività malevole ha portato alla luce la presenza di compagnie proprietarie di reti Internet o di gestori delle connessioni che, sotto l'influenza di organizzazioni criminali, od in ogni caso a conoscenza delle loro attività, ne tollerano la presenza. Spesso queste compagnie controllano network sufficientemente grandi da essere soggetti ad un range abbastanza ampio di tipologie di attività malevole, e non prendendo provvedimenti a riguardo permettono ai criminali informatici di avere una connessione sicura per perpetrare le loro attività illegali. Queste attività malevole si suddividono in diverse tipologie, che vanno dallo spam, al phishing, allo spazio per condividere software piratati o contenuti pedo-pornografici, e così via. In particolare il database di FIRE raccoglie dati inerenti alle seguenti tipologie di attività:

Spam Per spam si intende l'invio di grandi quantità di messaggi indesiderati, a migliaia di indirizzi mail contemporaneamente. Gli scopi di questa attività sono generalmente di tipo commerciale, ed il contenuto pubblicitario può essere di varia natura.

Phishing Il phishing è un'attività illegale che, sfruttando messaggi di posta elettronica fasulli o messaggi istantanei, mira a ottenere l'accesso a informazioni personali o riservate con la finalità del furto d'identità. Imitando nella grafica, nei loghi e nel contenuto le comunicazioni dei siti istituzionali, il phisher raggira l'utente portandolo a rivelare dati personali quali numero di conto corrente, numero di carta di credito, etc.

[3] B. Stone-Gross, C. Kruegel, K. Almeroth, A. Moser, and E. Kirda. Fire: Finding rogue networks. In Proc. of the 2009 Annual Comp. Security Applications Conference, ACSAC '09, pages 231-240, Washington, DC, USA, 2009. IEEE Comp. Society.

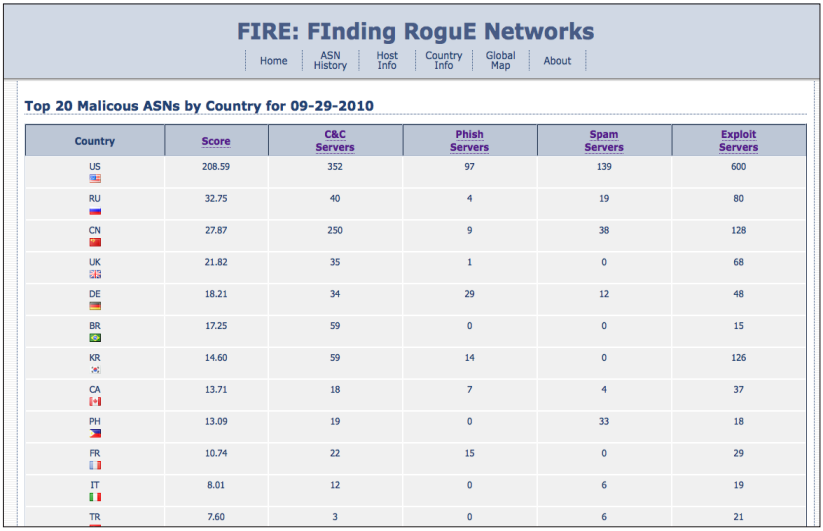
Malware Viene definito malware un qualsiasi software creato con lo scopo di recare danno. I malware possono agire recando danni diretti sul computer che viene infettato, lavorando in background, ad esempio trasformando il pc infettato in una botnet, o ancora carpire informazioni o password personali digitate sul pc infettato per poi rimandarle al mittente.

Botnet Una botnet è una rete di computer collegati ad internet che fanno parte di un insieme più grande controllato da un'unica entità, il botmaster. Si parla di botnet quando un computer viene infettato, consentendo di conseguenza al botmaster di controllarne il sistema da remoto. I controllori della botnet possono in questo modo sfruttare i sistemi compromessi per scagliare attacchi del tipo DDoS (Distributed Denial of Service) contro qualsiasi altro sistema in rete o compiere altre operazioni illecite.

Spesso, una volta identificati ed esposti i server malevoli, si registra una rapida cessazione dell'attività, o un de-peering (sconnessione) da parte dei gestori delle connessioni, obbligando i criminali informatici a spostare le proprie operazioni, e causandogli importanti perdite di tempo e di denaro.
Per fare un esempio pratico, un network che offriva connessione sicura ad attività malevole è la Russian Business Network (RBN), che finì sulle prime pagine verso la fine del 2007. Varie fonti affermarono che la RBN ospitava siti web, exploit e malware responsabili di un numero significativo di truffe online e phishing. Una volta esposta pubblicamente, la RBN cessò le sue attività in San Pietroburgo, ma solo per spostarsi e riprenderle poco più tardi su altri network. Più recentemente, un report ha segnalato Atrivio (Intercage), una compagnia stanziata negli USA, ospitante anche lei attività malevole. Poco dopo la scoperta di Atrivio, altri due network, conosciuti come McColo e 3FN(Triple Fiber Network), furono messi in luce per i loro legami con attività illecite. In entrambi i casi l'esposizione pubblica di queste relazioni ha portato i gestori di queste reti a rapidi provvedimenti per tagliare la connessione alle organizzazioni criminali.

Ovviamente queste reti non sono l'unico mezzo a disposizione di questa fiorente economia, responsabile di moltissimi problemi legati alla sicurezza degli utenti di Internet. Nel corso degli ultimi anni si è registrato un ampio aumento dell'uti-

lizzo di botnet per nascondere la fonte originaria degli attacchi. In ogni caso si è visto però come l'esposizione pubblica di questi network possa rivelarsi utile. Per prima cosa i criminali temono l'attenzione pubblica, come risultato di una crescente attenzione da parte dei media tutte i casi precedentemente descritti hanno visto una cessazione, o in ogni caso diminuzione, delle attività. Nonostante in molti casi la tendenza sia quella di spostare il proprio business su di un altro network, questa operazione può richiedere un arco tempo durante il quale le attività malevole devono necessariamente cessare. La rapida individuazione di queste attività potrebbe rendere difficile per i criminali informatici stabilire una base sicura per i loro traffici, obbligandoli a continui spostamenti.



FIRE: FInding RoguE Networks

Home | ASN History | Host Info | Country Info | Global Map | About

Top 20 Malicous ASNs by Country for 09-29-2010

Country	Score	C&C Servers	Phish Servers	Spam Servers	Exploit Servers
US	208.59	352	97	139	600
RU	32.75	40	4	19	80
CN	27.87	250	9	38	128
UK	21.82	35	1	0	68
DE	18.21	34	29	12	48
BR	17.25	59	0	0	15
KR	14.60	59	14	0	126
CA	13.71	18	7	4	37
PH	13.09	19	0	33	18
FR	10.74	22	15	0	29
IT	8.01	12	0	6	19
TR	7.60	3	0	6	21

▲ **fig 02 | Database FIRE**
Schermata del database di FIRE, tratta da una delle pagine del sito maliciousnetworks.org

3.2 Software e prodotti sul mercato

Anche uscendo dal campo della ricerca scientifica, si possono trovare un gran numero di prodotti e strumenti esistenti, creati sia da privati che da aziende, utili di volta in volta a raccontare o analizzare in maniera differente il fenomeno. Si tratta di sistemi sia gratuiti che a pagamento, sia alla portata di tutti che creati appositamente per le aziende. Molti di questi non sono veri e propri strumenti per l'analisi, ma si limitano a fornire un report sulle attività registrate. Nelle pagine che seguono si è fatta una piccola raccolta dei più interessanti trovati in rete, proponendone una suddivisione per tipologie.

3.2.1 Attività di report

Un gran numero di software e prodotti presenti sul mercato si configurano più come mezzi adatti a fornire report sul fenomeno, piuttosto che come veri e propri strumenti di analisi dello stesso. Si tratta infatti di prodotti, a volte anche interattivi, che tuttavia non permettono un lavoro di analisi e ricerca attiva da parte dell'utente, ma si limitano a presentargli la situazione come questa è stata registrata.

Conficker worm visualizations

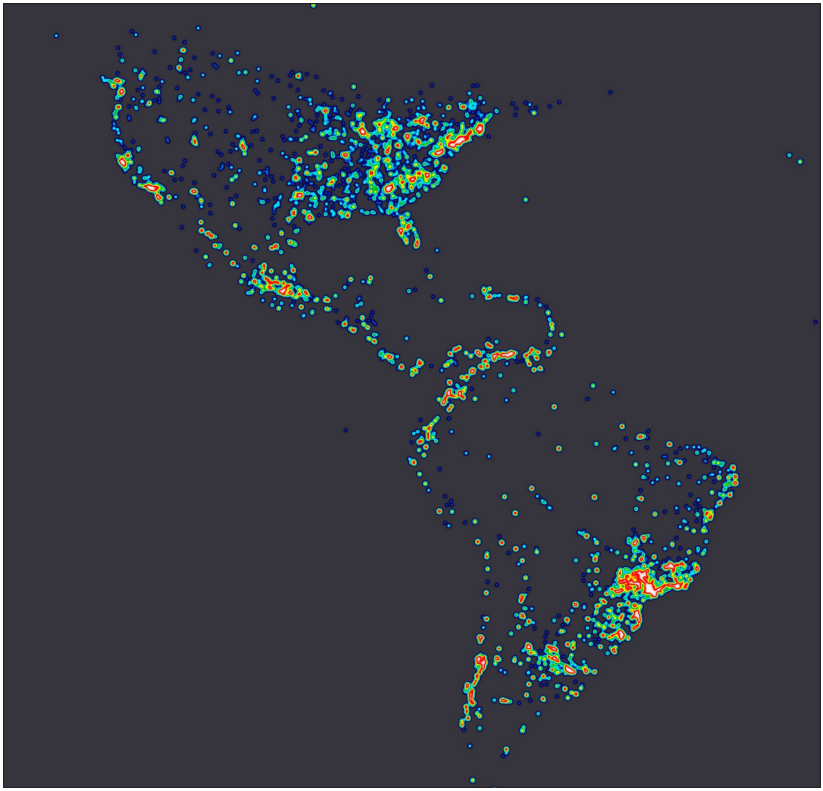
Autore Team Cymru
Anno gennaio 2009
Attività visualizzata Conficker Worm
Visualizzazione Statica georeferenziata

Visualizzazione degli IP infettati dal worm Conficker nella giornata del 29 gennaio 2009. Nonostante la mappa geografica sottostante sia stata rimossa, il numero degli attacchi è così elevato da ridisegnare il contorno dei continenti.

tasso di infezione

minore maggiore

▶ **fig 03a-b | Conficker Worm Visualizations**
Immagini tratte dal sito http://www.team-cymru.org/Monitoring/Malevolence/conficker.html



Internet malicious activity World Map

Autore Team Cymru
Anno -
Attività visualizzata Malicious Activity
Visualizzazione Dinamica georeferenziata

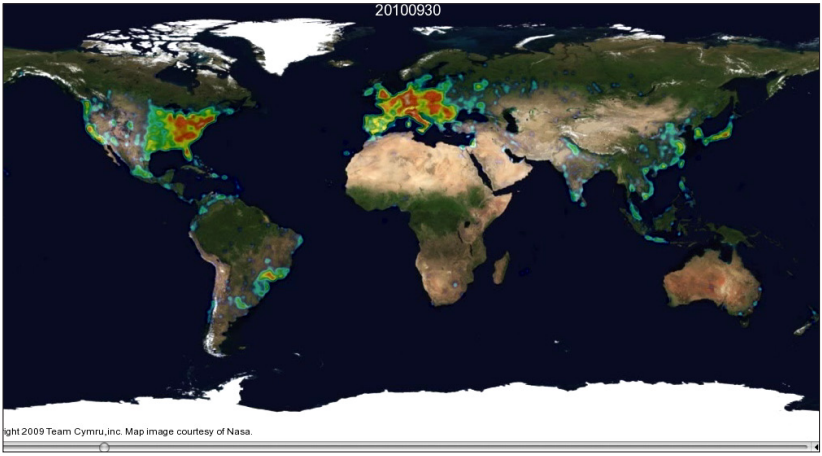
Aggiornamento giornaliero

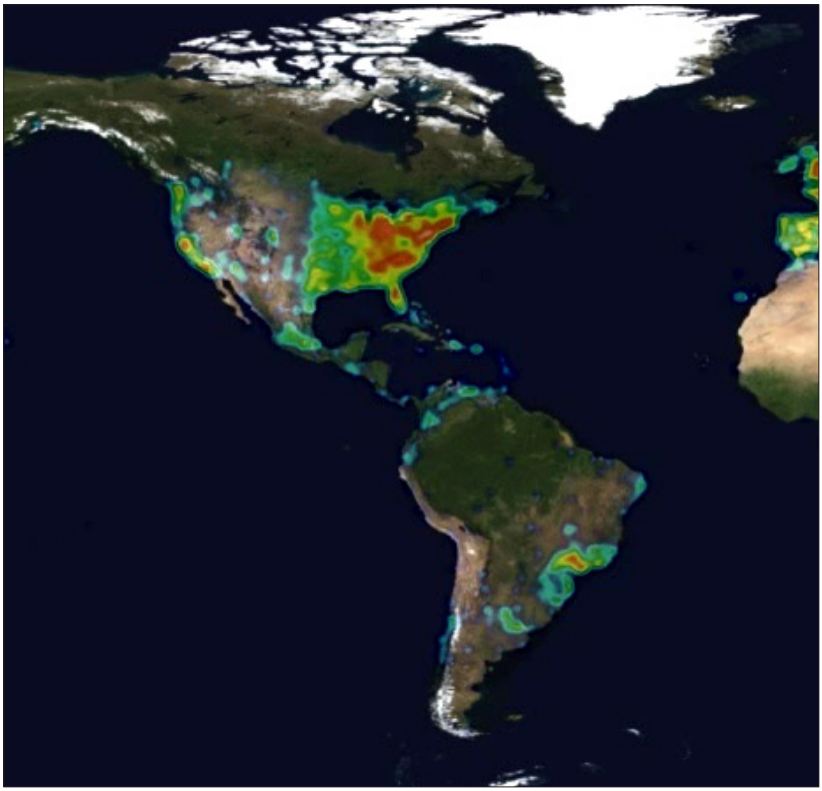
Mappa georeferenziata sulla quale viene posizionata l'attività malevola registrata sulla rete Internet nel corso dell'intera giornata precedente la visualizzazione. La mappa è fruibile sotto forma di video.

tasso di infezione

minore maggiore

▶ **fig 04a-b | Internet malicious activity World Map**
Immagini tratte dal sito http://www.team-cymru.org/Monitoring/Malevolence/maps.html


20100930
ight 2009 Team Cymru, inc. Map image courtesy of Nasa.



Internet malicious activity Hilbert Map

Autore Team Cymru
Anno -
Attività visualizzata Malicious Activity
Visualizzazione Statica

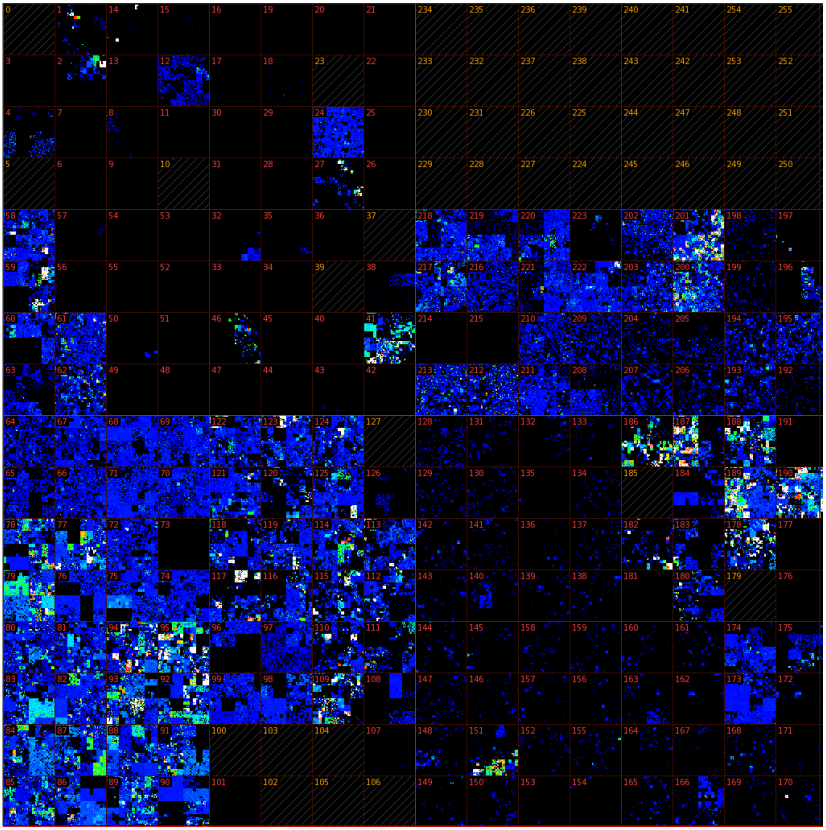
Aggiornamento giornaliero

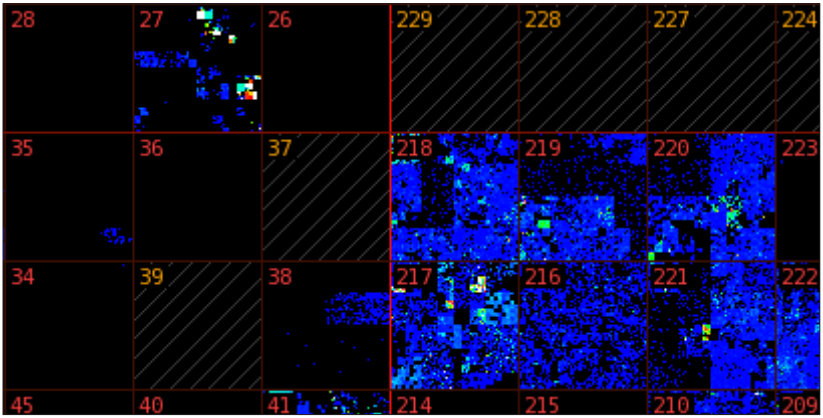
Mappa ad aggiornamento automatico giornaliero dell'attività malevola registrata in internet negli ultimi 30 giorni. Il numero visualizzato in ogni blocco rappresenta la prima parte dell'indirizzo IP (riconducibile ad una zona fisica del pianeta). I blocchi arancioni indicano lo spazio non ancora allocato.

tasso di infezione

minore maggiore

▶ **fig 05a-b | Internet malicious activity Hilbert Map**
Immagini tratte dal sito http://www.team-cymru.org/Monitoring/Malevolence/maps.html




28
27
26
229
228
227
224
35
36
37
218
219
220
223
34
39
38
217
216
221
222
45
40
41
214
215
210
209

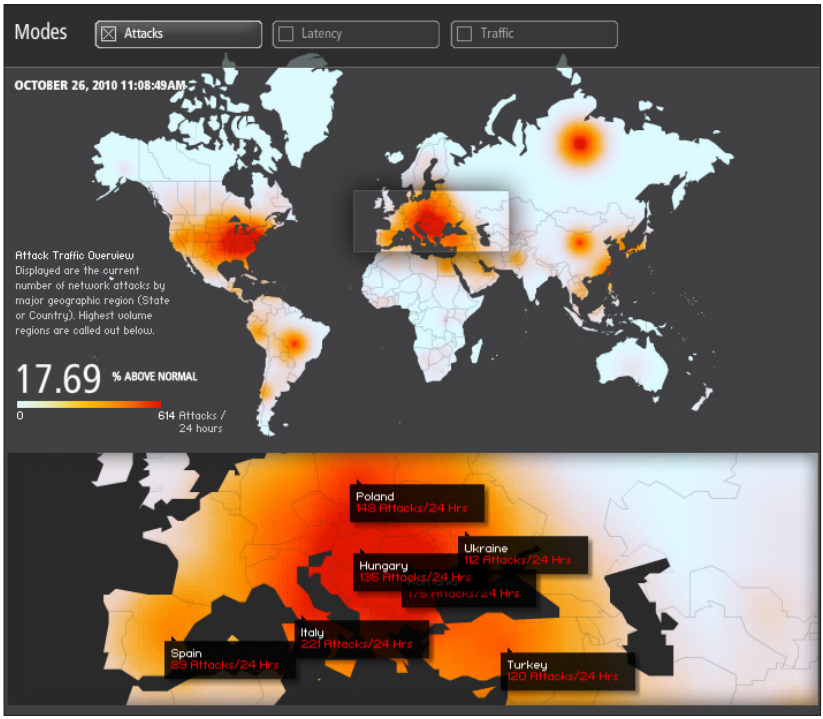
Akamai Real-time Web Monitor

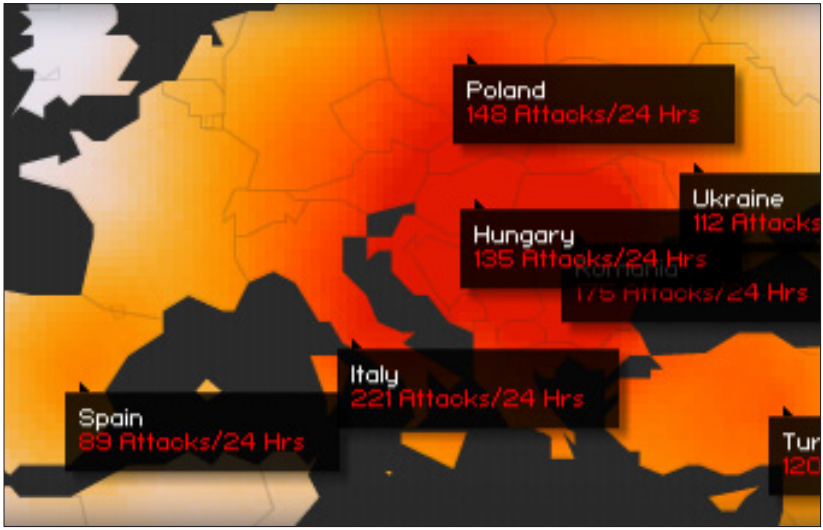
Autore Akamai
Anno 2007
Attività visualizzata Malicious Activity
Visualizzazione Dinamica georeferenziata interattiva

Live

Mappa georeferenziata sulla quale vengono visualizzati in tempo reale il numero di attacchi perpetrati nella giornata corrente, suddivisi per stati. Possibilità di visualizzare anche altri parametri (latency e traffic) sulla stessa mappa.

▶ **fig 06a-b | Akamai Real-time Web Monitor**
Immagini tratte dal sito http://www.akamai.com/html/technology/dataviz1.html


Modes
Attacks
Latency
Traffic
OCTOBER 26, 2010 11:08:49AM
Attack Traffic Overview
Displayed are the current number of network attacks by major geographic region (State or Country). Highest volume regions are called out below.
17.69
% ABOVE NORMAL
0
614 Attacks / 24 hours
Poland
148 Attacks/24 Hrs
Ukraine
112 Attacks/24 Hrs
Hungary
135 Attacks/24 Hrs
Italy
221 Attacks/24 Hrs
Spain
89 Attacks/24 Hrs
Turkey
120 Attacks/24 Hrs


Poland
148 Attacks/24 Hrs
Ukraine
112 Attacks
Hungary
135 Attacks/24 Hrs
Italy
221 Attacks/24 Hrs
Spain
89 Attacks/24 Hrs

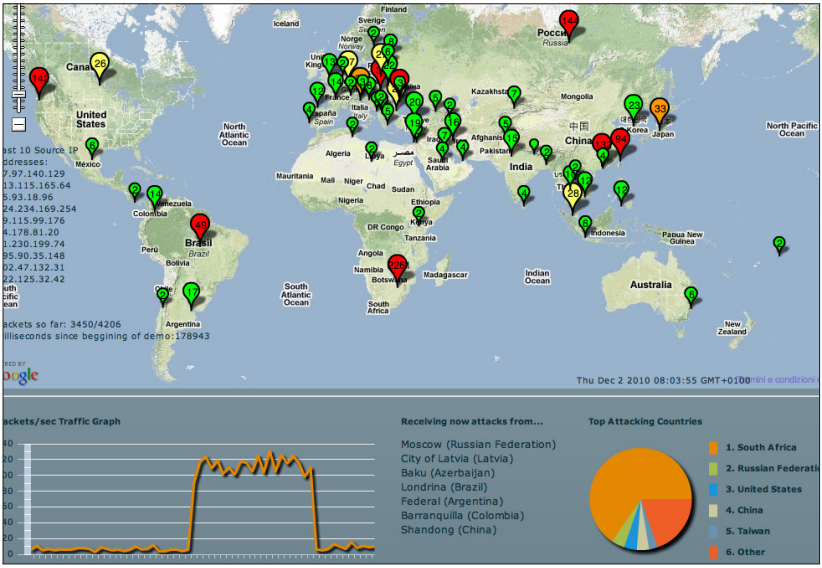
NoAH.honeypots TrGeo

Autore NoAH - Forth ICS - DCS Lab
Anno luglio 2008
Attività visualizzata Malicious Activity
Visualizzazione Dinamica georeferenziata interattiva

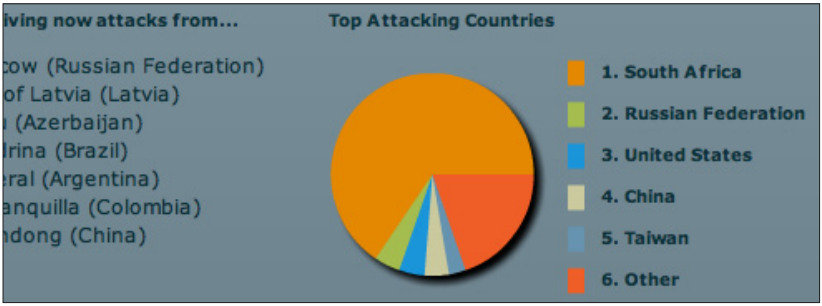
Live

Mappa georeferenziata che mostra in tempo reale da dove vengono effettuati gli attacchi nel mondo. Si basa sul principio degli Honeypots ed oltre alla mappa, nella parte inferiore, offre anche due grafici riassuntivi.

▶ **fig 07a-b-c | NoAH.honeypots TrGeo**
Immagini tratte dal sito https://stats.fp6-noah.org/trgeo.php


Receiving now attacks from...
Moscow (Russian Federation)
City of Latvia (Latvia)
Baku (Azerbaijan)
Londrina (Brazil)
Federal (Argentina)
Barranquilla (Colombia)
Shandong (China)
Top Attacking Countries
1. South Africa
2. Russian Federati
3. United States
4. China
5. Taiwan
6. Other
Thu Dec 2 2010 08:03:55 GMT+0100


Russia
Kazakhstan
Mongolia
China
India
Japan
Korea
Algeria
Libya
Egypt
Saudi Arabia
Pakistan
Afghanis
Mali
Niger
Chad
Sudan
Nigeria
Ethiopia
France
Italia
Italy
España
Spain
Norway

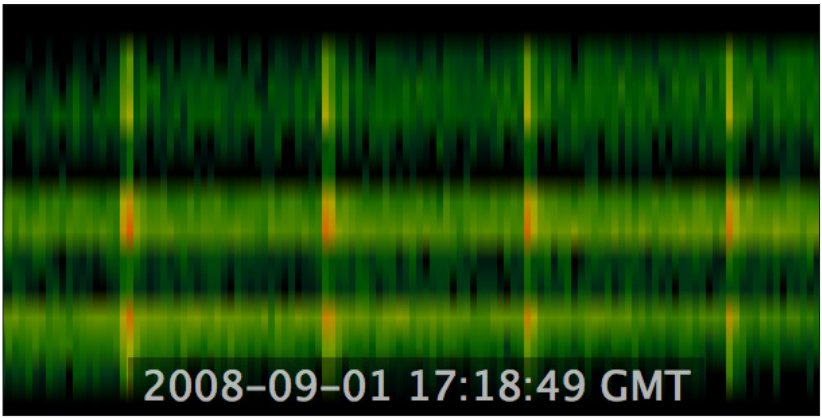

Top Attacking Countries
1. South Africa
2. Russian Federation
3. United States
4. China
5. Taiwan
6. Other

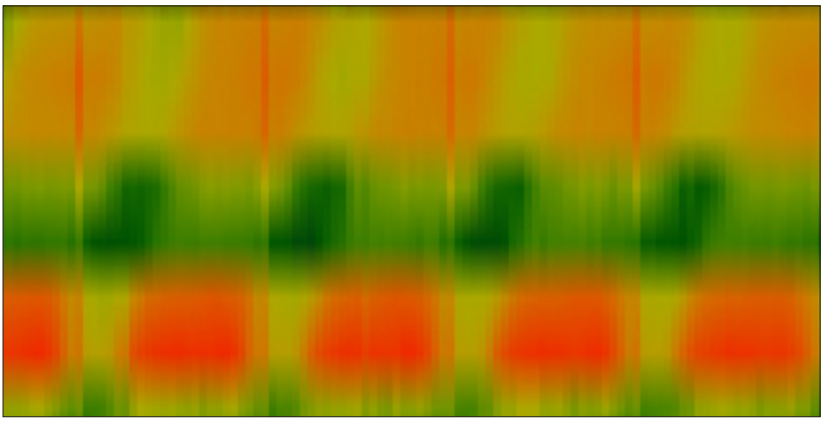
Spectral view on activity

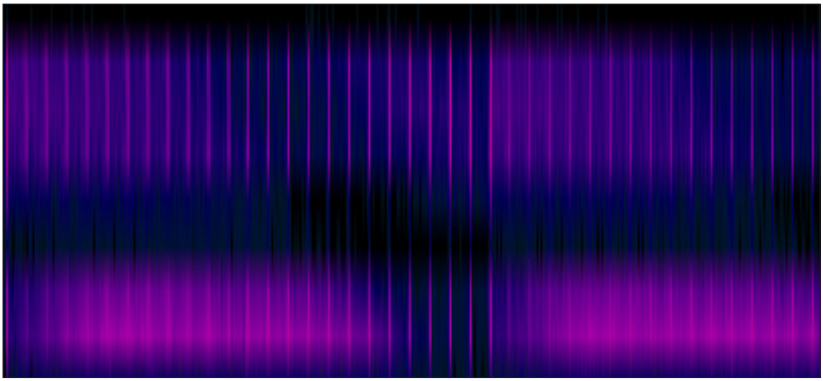
Autore Clarified networks
Anno -
Attività visualizzata Bot
Visualizzazione Spettro di attività

Visualizzazione delle botnet tramite uno spettro di attività. Sull'asse delle y vi è la latitudine, su quello delle x il tempo, il colore indica l'intensità dell'attività stessa. Attraverso queste visualizzazioni viene mostrato come esistano degli andamenti periodici dell'attività osservata nel tempo.

▶ **fig 08a-b-c | Spectral view on activity**
Immagini tratte dal sito
https://www.clarifiednetworks.com/ClarifiedVisualizationGallery#Situation_Rooms_-_Intuitive_views


2008-09-01 17:18:49 GMT





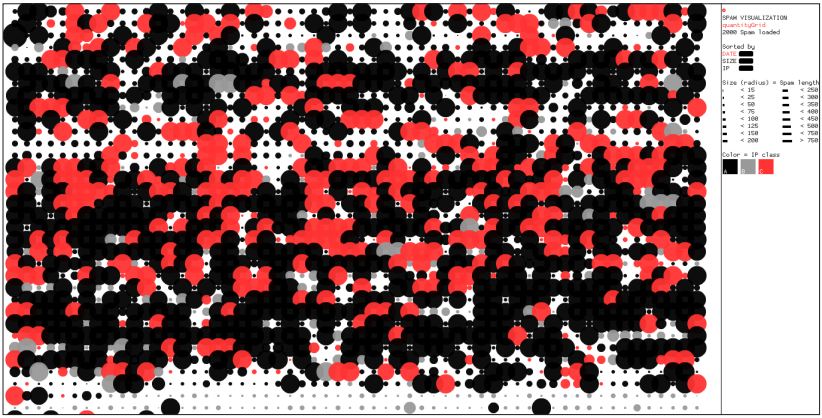
Spam Visualization

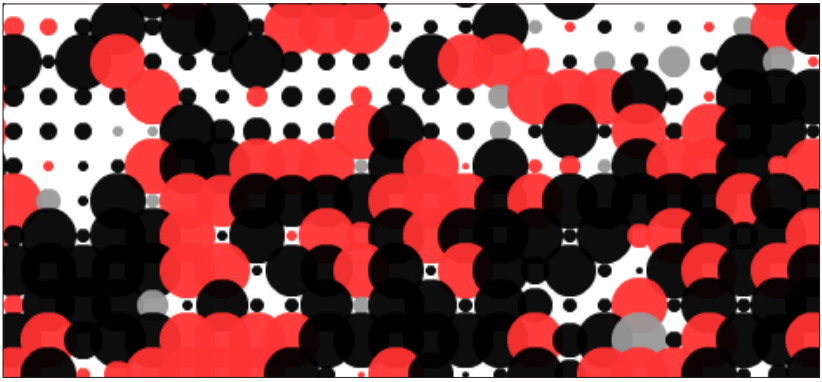
Autore Kim Asendorf - kaubonschen creative studio
Anno 2009
Attività visualizzata Spam
Visualizzazione Diverse visualizzazioni interattive

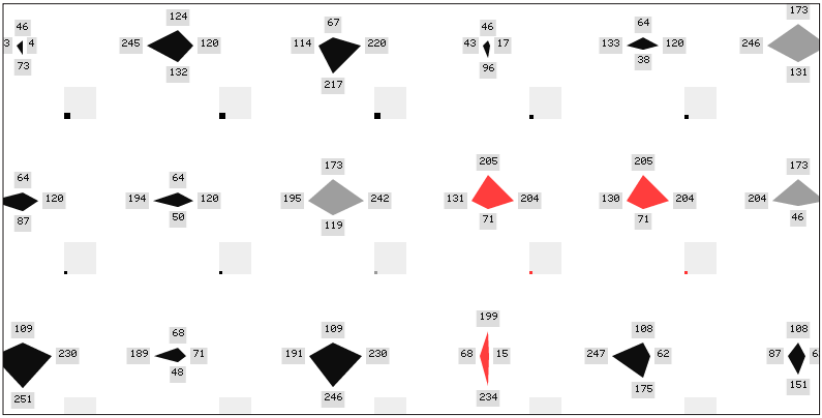
Live

Strumento che analizza e visualizza, sotto forma di diversi diagrammi, un database chiamato Spam Catalog, creato sempre da Kim Asendorf. Il linguaggio grafico, a detta dell'autore, tende ad essere il più semplice possibile.

▶ **fig 09a-b-c | Spam Visualization**
Immagini tratte dal sito http://spamvisualization.net/


SPAM VISUALIZATION
quantityGrid
20000 Spam Loaded
Sorted by
DATE
SIZE
IP
Size (radius) = Spam length
Color = IP class




46 3 4 73
124 245 120 132
67 114 220 217
46 43 17 96
64 133 120 38
173 246 131
64 120 87
64 194 120 50
173 195 242 119
205 131 204 71
205 130 204 71
173 204 46
109 230 251
68 189 71 48
109 191 230 246
199 68 15 234
108 247 62 175
108 87 6 151

3.2.2 Controllare il proprio network

Tendenza soprattutto dei software a pagamento è quella di fornire dei prodotti in grado di controllare la situazione di sicurezza in cui si inserisce il proprio network, privato o lavorativo, al fine di aumentare la sicurezza personale dei propri dati.

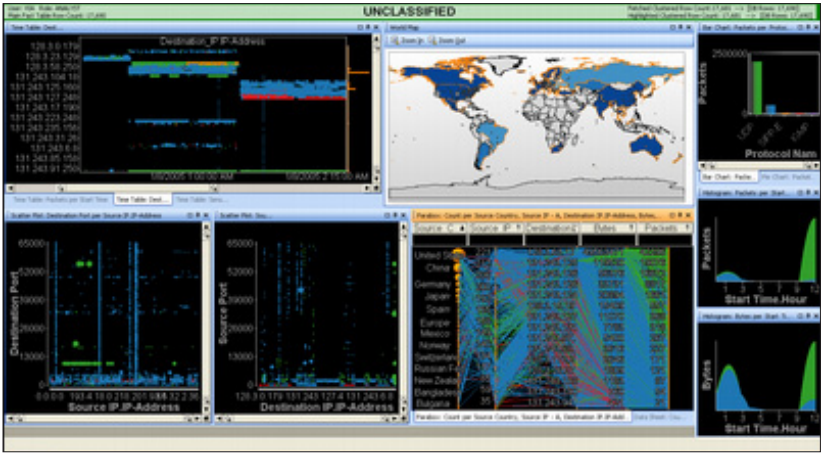
VIAssist

Autore Secure Decisions
Anno 2006
Attività visualizzata Network Traffic
Visualizzazione Varie tipologie di grafici e visualizzazioni

Live
Analisi locale
Software a pagamento

Software interattivo di analisi e visualizzazione del traffico locale legato al proprio network. Sviluppato al fine di ottenere un maggior controllo sulla sicurezza della propria rete, fa uso di diverse tipologie di grafici e visualizzazioni.

▶ fig 10a-b-c | VIAssist
Immagini tratte dal sito http://www.securedecisions.com/viassist


UNCLASSIFIED
Destination IP IP-Address
Protocol Name
Packets
Start Time.Hour
Bytes
Source IP IP-Address
Destination IP IP-Address
Destination Port
Source Port

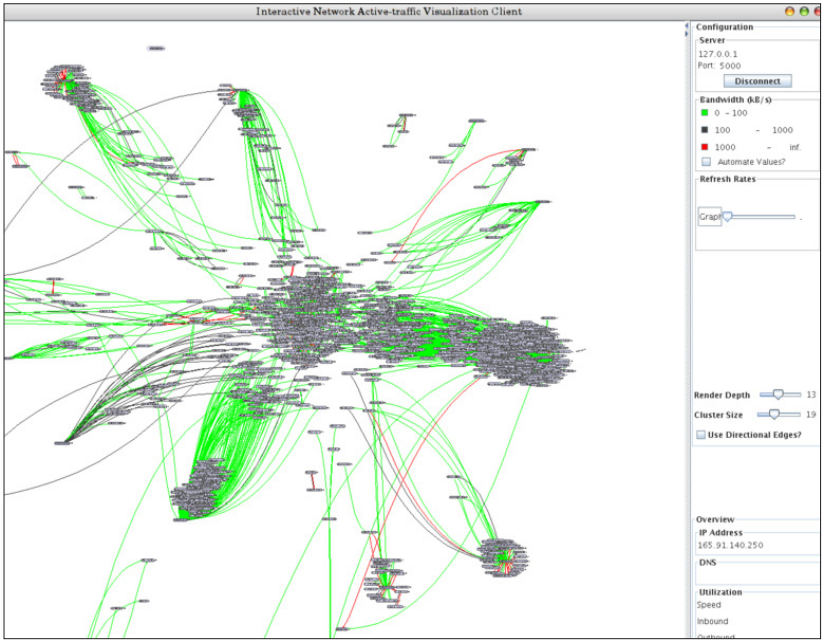
Interactive Network Active-traffic Visualization

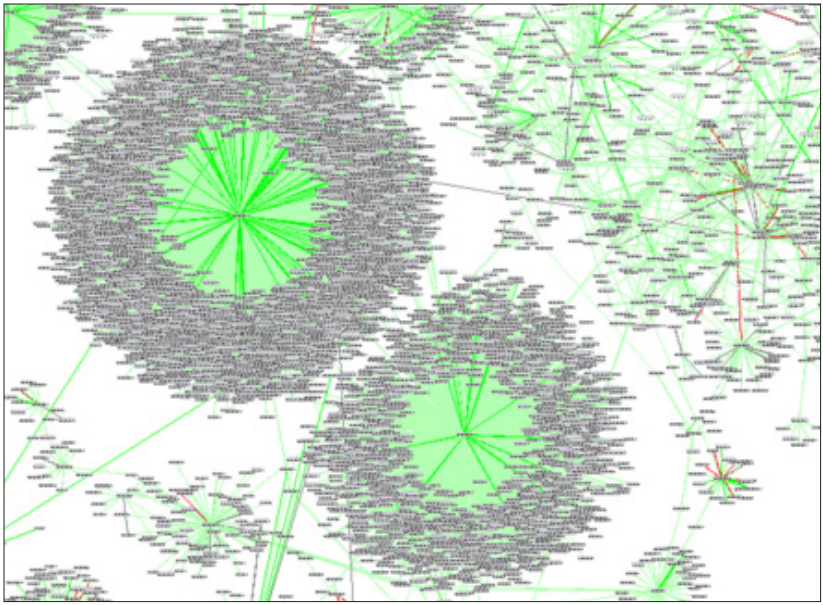
Autore Jeff Scaparra - Nathan Robinson
Anno 2007
Attività visualizzata Network Connections
Visualizzazione Grafi

Live
Analisi locale

Software interattivo in grado di visualizzare in tempo reale l'ambiente network che opera intorno alla macchina dal quale viene fatto girare. Attraverso le richieste e le comunicazioni che avvengono tra le varie macchine, costruisce dei grafi del network con cui si sta interagendo.

▶ **fig 11a-b | Interactive Network Active-traffic Visualization**
Immagini tratte dal sito http://inav.scaparra.com/


Interactive Network Active-traffic Visualization Client
Configuration
Server
127.0.0.1
Port: 5000
Disconnect
Bandwidth (kB/s)
0 - 100
100 - 1000
1000 - inf.
Automate Values?
Refresh Rates
Graph
Render Depth
13
Cluster Size
19
Use Directional Edges?
Overview
IP Address
165.91.140.250
DNS
Utilization
Speed
Inbound



Norton Internet Security 2011

Autore Symantec
Anno 2010
Attività visualizzata Malicious Activity
Visualizzazione Dinamica georeferenziata interattiva

Aggiornamento periodico
Analisi globale
Software a pagamento

Mappa georeferenziata situata all'interno di una delle schermate di configurazione del software antivirus. Con un aggiornamento ogni tot minuti vengono visualizzate sulla mappa le principali minacce rilevate. Passando col mouse sui punti luminosi (le minacce) si possono ottenere maggiori informazioni.

▶ **fig 12a-b | Norton Internet Security 2011**
Immagini tratte dal software Norton Internet Security 2011


Norton Internet Security
System Status: Secure
Settings
Performance
Feedback
Account
Support
Computer Protection
Scan Now
History
Quarantine
Application Ratings
Run LiveUpdate
1 minute ago
Insight Protection
Details
Antivirus
Antispyware
SONAR Protection
Network Protection
Vulnerability Protection
Network Security Map
Smart Firewall
Intrusion Prevention
Email Protection
Web Protection
Logins
Cards
Parental Controls
Identity Safe
Browser Protection
Safe Surfing
Download Intelligence
Norton Protected You
Cities with the highest number of threats in: North America
9/6/2010
05:45 GMT
Map Details
TOP CITY THREATS 145 threats..... Toronto: 7877 threats..... Los Angeles: 7827 threats..... Houston: 6461 threats..... Chicago: 5950 threats..
Norton
SUBSCRIPTION STATUS:
366 Days Remaining
Renew
Activity Map
Online Family
Online Backup
Safe Web


lo: 7877 threats..... Los Angeles: 7827 threats..... Houston: 6461 threats

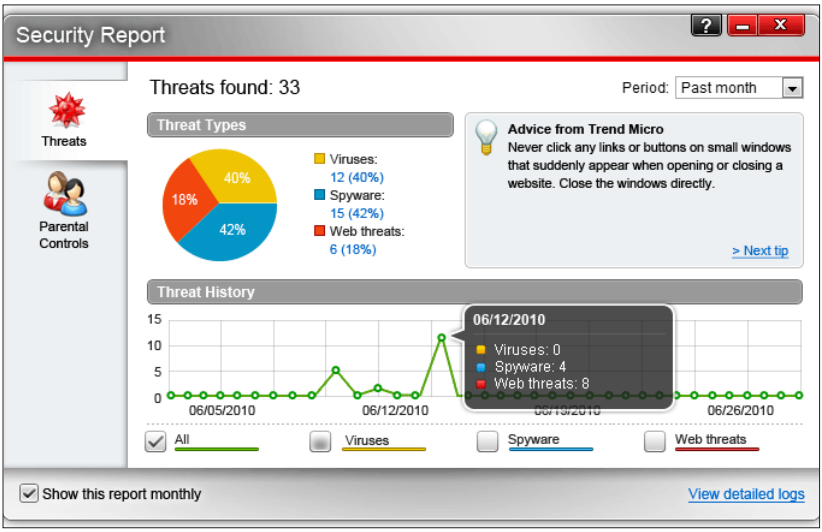
Titanium Internet Security 2011

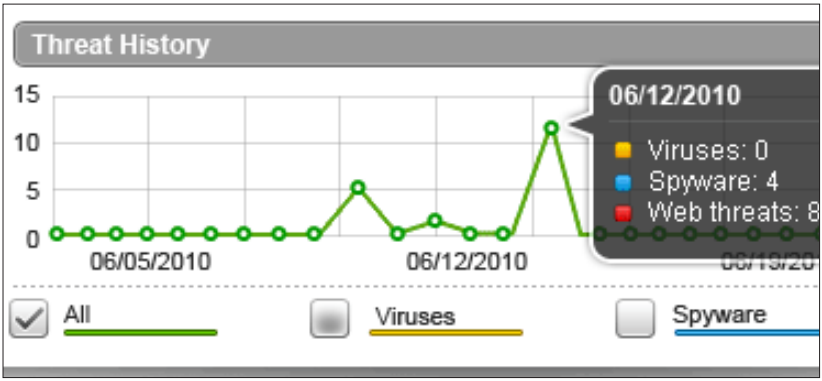
Autore Trend Micro
Anno 2010
Attività visualizzata Malicious Activity
Visualizzazione Dashboard interattiva

Aggiornamento periodico
Analisi locale
Software a pagamento

Dashboard del software antivirus dove attraverso dei grafici interattivi si possono andare a visualizzare i dati relativi al proprio network. L'utilizzo di visualizzazioni da parte dei software antivirus è un fenomeno in crescita.

▶ **fig 13a-b | Titanium Internet Security 2011**
Immagini tratte dal sito http://us.trendmicro.com/us/products/personal/internet-security/


Security Report
Threats found: 33
Period: Past month
Threats
Parental Controls
Threat Types
40%
18%
42%
Viruses: 12 (40%)
Spyware: 15 (42%)
Web threats: 6 (18%)
Advice from Trend Micro
Never click any links or buttons on small windows that suddenly appear when opening or closing a website. Close the windows directly.
> Next tip
Threat History
15
10
5
0
06/12/2010
Viruses: 0
Spyware: 4
Web threats: 8
06/05/2010
06/12/2010
06/26/2010
All
Viruses
Spyware
Web threats
Show this report monthly
View detailed logs


Threat History
15
10
5
0
06/12/2010
Viruses: 0
Spyware: 4
Web threats: 8
06/05/2010
06/12/2010
All
Viruses
Spyware

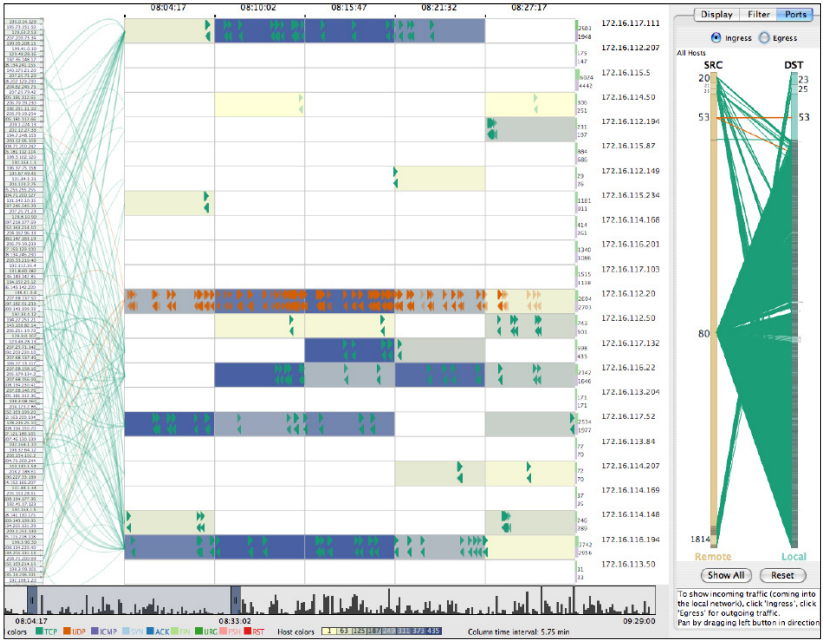
TNV: COMPUTER NETWORK TRAFFIC VISUALIZATION TOOL

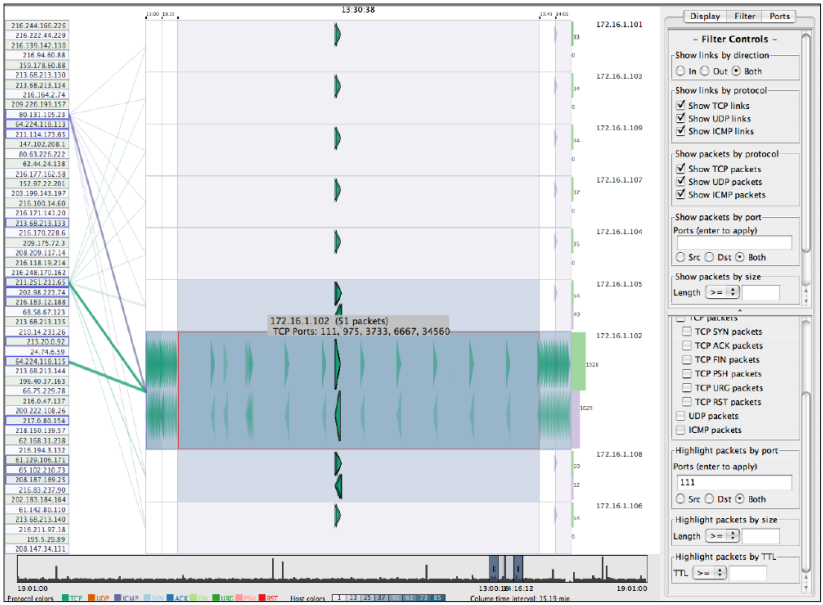
Autore Secure Decisions
Anno -
Attività visualizzata Network Traffic Analysis
Visualizzazione Varie tipologie di grafici e visualizzazioni

Live
Analisi locale

Software interattivo di visualizzazione per l'analisi del traffico legato ad un network locale. Tramite la visualizzazione si facilita la distinzione tra la normale attività di traffico e quella invece anormale e quindi potenzialmente legata ad una qualche tipologia di minaccia.

▶ **fig 14a-b | tnv: computer network traffic visualization tool**
Immagini tratte dal sito http://tnv.sourceforge.net/index.php


08:04:17
08:10:02
08:15:47
08:21:32
08:27:17
172.16.117.111
172.16.112.207
172.16.115.5
172.16.114.50
172.16.112.194
172.16.115.87
172.16.112.149
172.16.115.234
172.16.114.168
172.16.116.201
172.16.117.103
172.16.112.20
172.16.112.50
172.16.117.132
172.16.115.22
172.16.113.204
172.16.117.52
172.16.113.84
172.16.114.207
172.16.114.169
172.16.114.148
172.16.116.194
172.16.113.50
Display
Filter
Ports
Ingress
Egress
All Hosts
SRC
DST
20
53
80
1814
23
25
53
Remote
Local
Show All
Reset
To show incoming traffic (coming into the local network), click 'Ingress', click 'Egress' for outgoing traffic.
Pan by dragging left button in direction
08:04:17
08:33:02
09:29:09
colors
TCP
UDP
ICMP
SYN
ACK
FIN
URG
PSH
RST
Host colors
Column time interval: 5.25 min


13:30:38
216.244.166.226
216.222.44.219
216.139.142.133
216.94.60.88
159.178.90.88
213.68.213.130
213.68.213.134
216.164.2.74
209.226.193.157
80.121.105.23
64.224.118.113
211.114.173.65
147.102.208.1
80.63.226.222
62.44.24.138
216.127.116.58
152.97.32.201
203.196.143.197
216.100.14.60
216.171.147.20
213.68.213.133
216.170.228.6
209.175.72.3
208.209.117.14
216.118.16.214
216.248.170.162
211.251.211.65
202.98.221.74
216.183.16.188
64.58.67.123
213.68.213.135
210.14.221.26
213.25.0.52
24.74.6.59
64.224.118.115
213.68.213.144
196.40.37.183
66.75.229.78
216.0.47.137
200.222.118.26
217.0.80.154
216.150.139.57
62.108.31.218
216.194.1.112
61.129.106.171
61.104.216.73
208.187.186.25
216.83.237.90
202.183.184.164
61.142.80.170
213.68.213.140
216.211.97.18
195.5.25.89
208.147.34.131
172.16.1.101
172.16.1.103
172.16.1.109
172.16.1.107
172.16.1.104
172.16.1.105
172.16.1.102
172.16.1.108
172.16.1.106
172.16.1.102 (51 packets)
TCP Ports: 111, 975, 3733, 6667, 34560
Display
Filter
Ports
Filter Controls
Show links by direction
In
Out
Both
Show links by protocol
Show TCP links
Show UDP links
Show ICMP links
Show packets by protocol
Show TCP packets
Show UDP packets
Show ICMP packets
Show packets by port
Ports (enter to apply)
Src
Dst
Both
Show packets by size
Length
>=
TCP packets
TCP SYN packets
TCP ACK packets
TCP FIN packets
TCP PSH packets
TCP URG packets
TCP RST packets
UDP packets
ICMP packets
Highlight packets by port
Ports (enter to apply)
111
Src
Dst
Both
Highlight packets by size
Length
>=
Highlight packets by TTL
TTL
>=
19:01:00
13:00:18
16:12
19:01:00
Protocol colors
TCP
UDP
ICMP
SYN
ACK
FIN
URG
PSH
RST
Host colors
Column time interval: 15.19 min

3.2.3 Raccontare il fenomeno per sensibilizzare

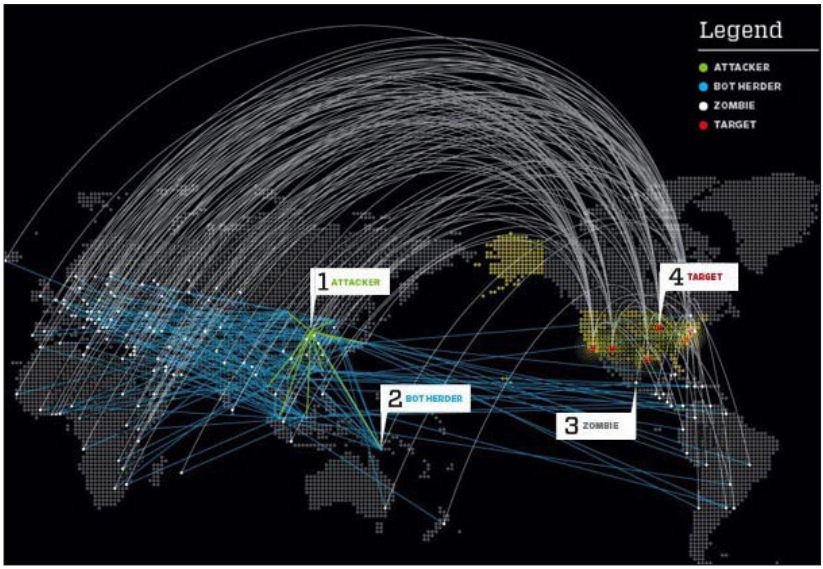
Un'altra tipologia di prodotti è quella creata al fine di raccontare, spiegare il fenomeno all'utente per informarlo e sensibilizzarlo verso un mondo a lui ancora non familiare. É questo l'esempio di alcune infografiche statiche apparse su magazine e riviste, che hanno come obiettivo proprio quello di spiegare il funzionamento di alcuni fenomeni legati alle minacce informatiche ad un pubblico non del settore, o anche campagne e prodotti web based che hanno come fine quello di informare l'utente circa i pericoli che si possono incontrare sulla rete.

When Bots Attack

Autore Catalogtree and Systemantics
Anno Pubblicazione settembre 2007
Attività visualizzata Bot
Visualizzazione Statica georeferenziata

Questa infografica statica, uscita sul numero 15.09 della rivista Wired america, presenta in maniera semplice ma estremamente comprensibile il funzionamento di un attacco botnet. Pur trattandosi di un esempio, e quindi non basato su dati reali, e pur essendo una semplice tavola, e quindi non strumento interattivo a supporto di un processo di analisi o di ricerca, questo lavoro è in ogni caso di interesse in quanto tenta, attraverso una narrazione grafica, di spiegare ed avvicinare un vasto pubblico, presumibilmente poco esperto del settore, ad una problematica di estrema attualità.

▶ **fig 15a-b | When Bots Attack**
Immagini tratte dal sito http://www.wired.com/politics/security/magazine/15-09/ff_estonia_bots


Legend
ATTACKER
BOT HERDER
ZOMBIE
TARGET
1 ATTACKER
2 BOT HERDER
3 ZOMBIE
4 TARGET


1 ATTACKER
2 BOT HERDER
3 ZOMBIE

Norton Cybercrime Index

Autore Symantec
Anno Febbraio 2011
Attività visualizzata Cybercrime
Visualizzazione Serie di visualizzazioni statiche interattive

Aggiornamento giornaliero

Presentato a Londra nel febbraio 2011, il Norton Cybercrime Index di Symantec nasce come un tool gratuito a disposizione di tutti coloro che giornalmente navigano su internet, che si pone come obiettivo quello di fornire in tempo reale un bollettino dello stato sicurezza della rete al fine di sensibilizzare gli utenti su un fenomeno ancora largamente sottovalutato. In particolar modo visualizza dati relativi a minacce quali malware, spam, phishing e furto d'identità.

▶ **fig 16a-b | Norton Cybercrime Index**
Immagini tratte dal sito http://it.norton.com/theme.jsp?themeid=protect_yourself


CYBERCRIME
INDEX
Norton
by Symantec
CYBERCRIME
INDEX
63
index
3%
change
PROTECT
YOURSELF


HISTORIC TREND CHART
360
180
0
Identity Theft Index: 262 1.95%
Cyber Crime Index: 143 1.42%
Fraud Index: 99 -6.60%
Malware Index: 61 -3.17%
Spam Index: 22 -8.33%
JAN '11
FEB '11
MAR '11
APR '11
Click and drag mouse to zoom into a specific date. Double click on the ch
Data used in ID Theft and Fraud index from ID Analytics®

3.2.4 La visualizzazione come processo artistico

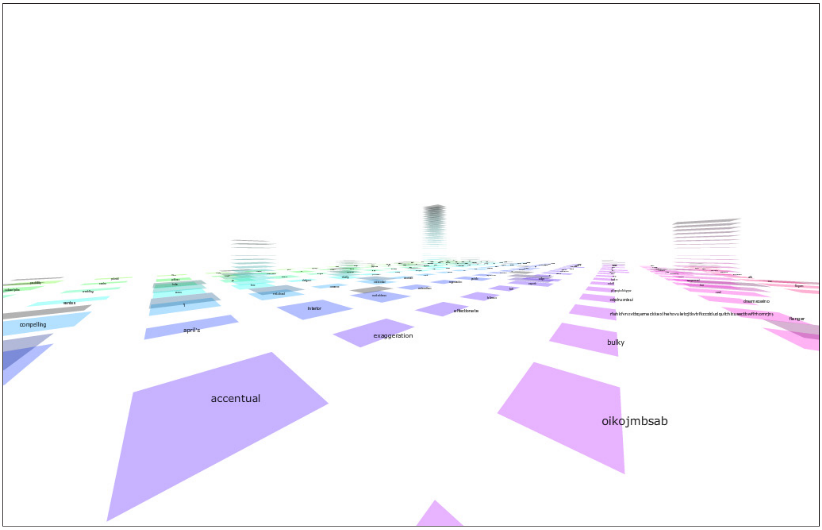
In altri casi ancora la visualizzazione di dati inerenti le minacce informatiche tende a sfociare verso quello che appare più come un processo artistico, rispetto a quello che potrebbe essere un progetto di comunicazione di informazione.

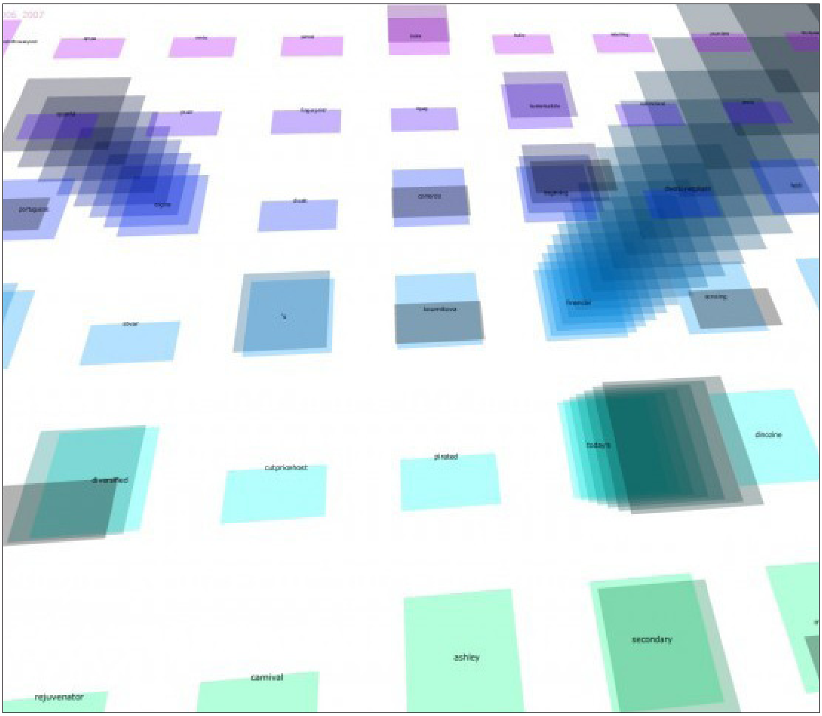
SPAMOLOGY

Autore Irad Lee
Anno 2007
Attività visualizzata Spam
Visualizzazione Rappresentazione audiovisiva

Visualizzazione della frequenza delle parole contenute nelle e-mail di spam. La visualizzazione si basa sull'analisi di un archivio privato che contiene informazioni sulle e-mail di spam dal 1998 al 2007, collezionate in varie parti del mondo. I dati sono visualizzati in uno spazio tridimensionale dove le parole più popolari sono rappresentate come strutture rettangolari di varie altezze. L'altezza è determinata dal numero di occorrenze di una stessa parola nello stesso anno.

▶ **fig 17a-b | Spamology**
Immagini tratte dal sito http://www.iradlee.com/projects/spamology/


accentual
oikojmbsab
bulky
exaggeration
april's
compelling

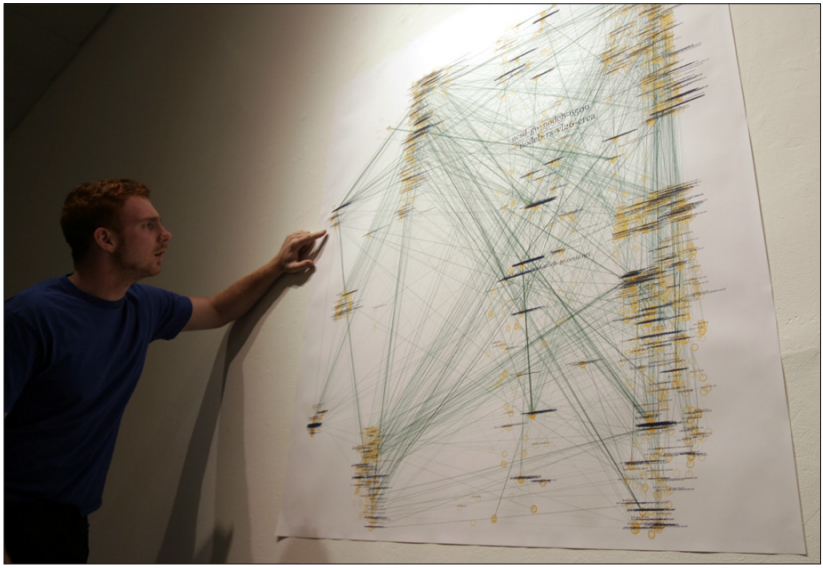

rejuvenator
carnival
ashley
secondary
diversified
today's

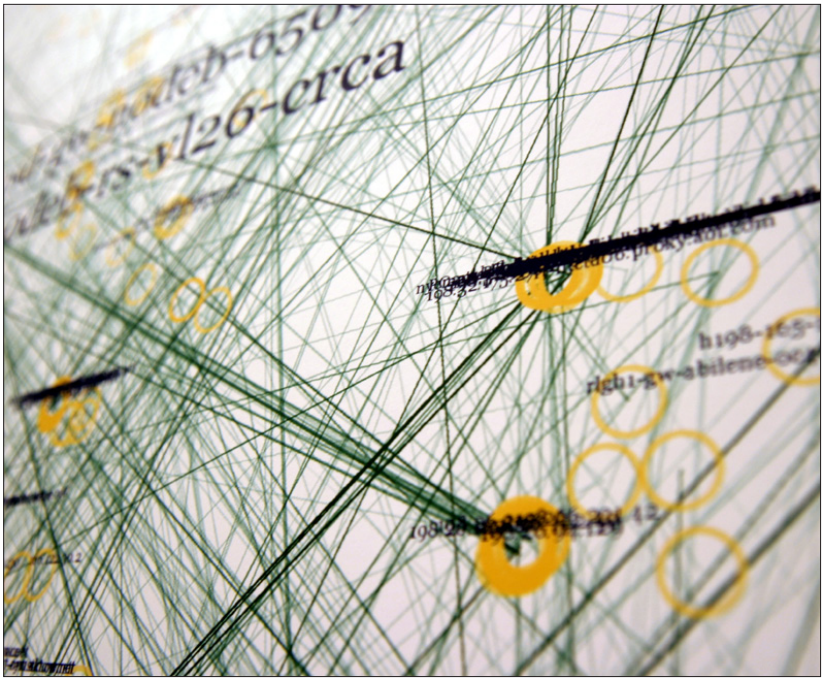
Respam

Autore Alex Dragulescu
Anno -
Attività visualizzata Spam
Visualizzazione Grafo Statico

Visualizzazione riconducibile al mondo della net.art, che tuttavia restituisce una mappa di un'attività di spam legata ad alcuni indirizzi email sotto forma di grafo. I grafi sono una tipologia di visualizzazione abbastanza usata nel campo della network visualization.

▶ **fig 18a-b | Respam**
Immagini tratte dal sito http://sq.ro/respam.php




rigb1-sw-crca

3.3 Come viene visualizzato il crimine nel mondo fisico

Nonostante fino a questo si sia parlato di minacce e criminalità dal punto di vista prettamente informatico, per avere un quadro generale completo dei diversi metodi di visualizzazione presenti sul mercato in ambito di gestione della criminalità, si rende necessario fare anche cenno a quelle visualizzazioni che si occupano di visualizzare il crimine nel mondo fisico. In questo ambito nel corso della storia sono stati fatti davvero un gran numero di lavori e sperimentazioni, ed il campo di indagine è molto ampio e sfaccettato. Qui di seguito sono stati riportati alcuni esempi scelti tra quelli che, per modalità di gestione dei dati e visualizzazione, sono risultati maggiormente interessanti per la nostra ricerca.

Oakland e San Francisco Crimespotting

La Oakland e la San Francisco Crimespotting map, sono due mappe interattive, sviluppate nella medesima maniera, che si propongono come strumento per la visualizzazione e la comprensione della criminalità a livello cittadino. Basate su di un database aggiornato giornalmente, si compongono di una mappa dettegliata della città in questione dove i diversi crimini segnalati vengono posizionati attraverso dei piccoli pallini colorati. Il colore dei pallini varia in base alla tipologia di crimine registrata e cliccando su di essi un popup descrive nel dettaglio ciò che è stato segnalato in quel determinato punto. É presente anche una timeline che permette di modificare il periodo di tempo che si vuole prendere in esame.

Murder: New York City. Homicides 2003-2009

Similmente alle due mappe precedenti, anche in questo caso la visualizzazione è composta da una mappa geografica della città di new york, dove tramite dei piccoli pallini colorati vengono posizionati gli omicidi perpetrati nella città. Attraverso un selettore sulla sinistra è possibile modificare la visualizzazione in base ad alcuni parametri prestabiliti, mentre una timeline posizionata nella parte alta della stessa permette di scegliere l'anno di cui si stanno andando a visualizzare i dati.

The truth about Crime: The Oxford Crime Map

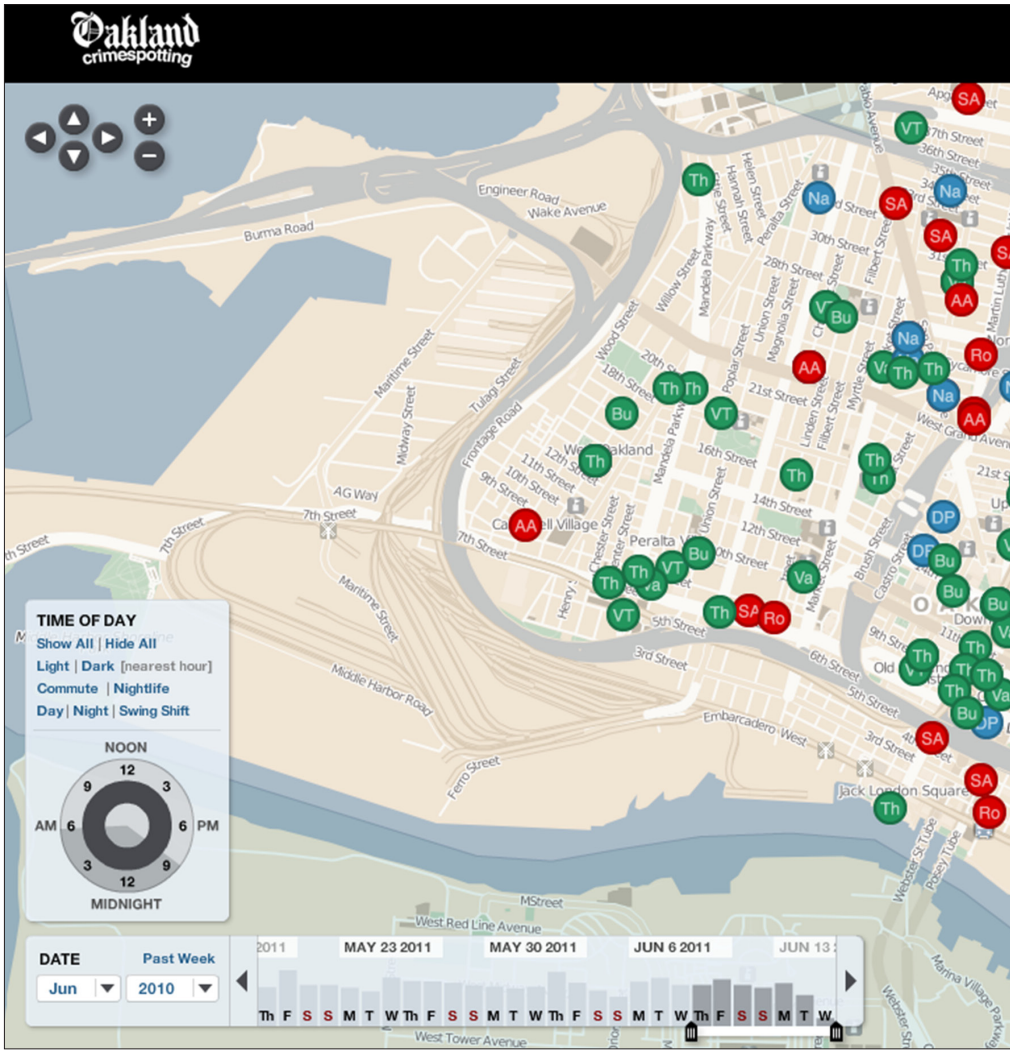
Anche in questo caso la visualizzazione è composta da una mappa geografica, questa volta della città di Oxford, dove però inizialmente non è riportato alcun dato.

Tramite l'utilizzo di alcuni menù a tendina posti sulla sinistra, l'utente può decidere quali dati andare a visualizzare sulla mappa, sovrapponendo in questo modo alla mappa cittadina una heat map la cui colorazione varia in base alla quantità di crimini registrati in ogni zona.

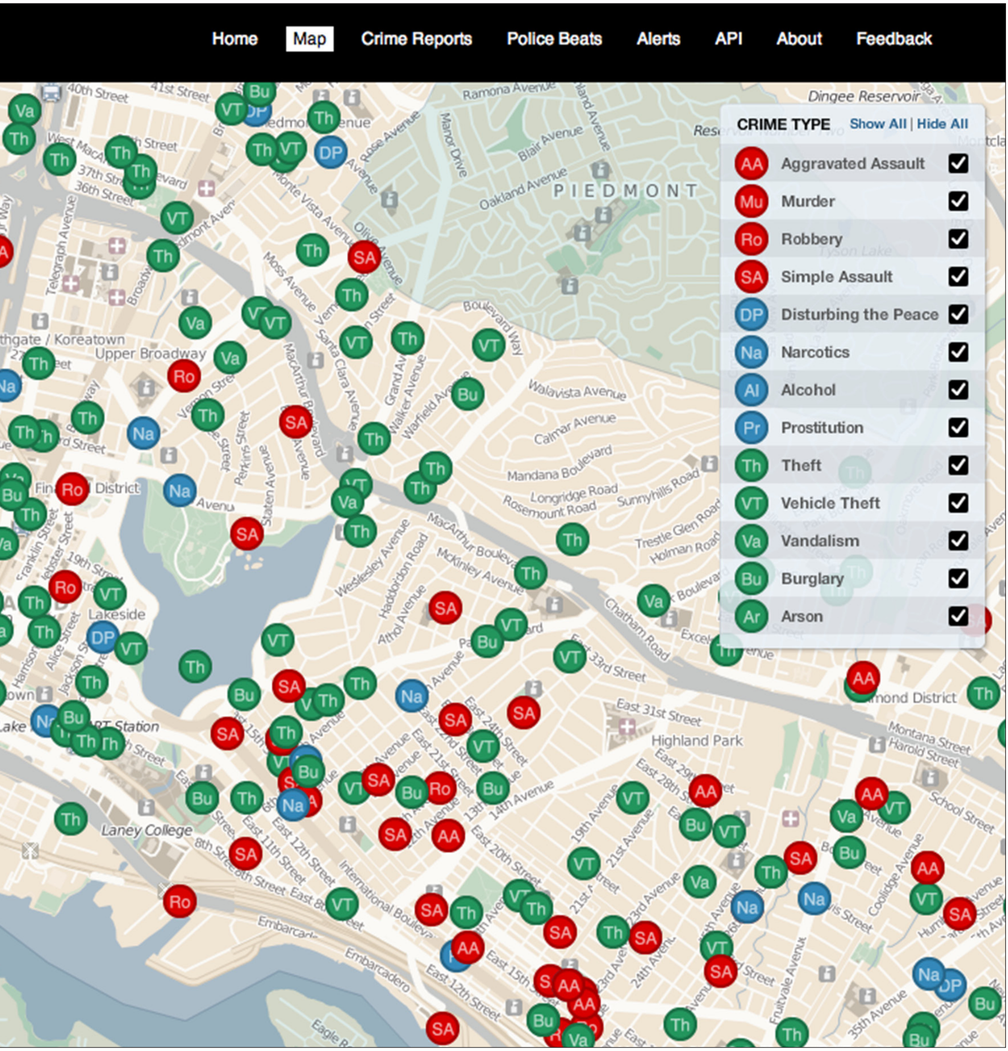
▶ fig 19 | Oakland Crimespotting
Immagine tratta dal sito http://oakland.crimespotting.org/

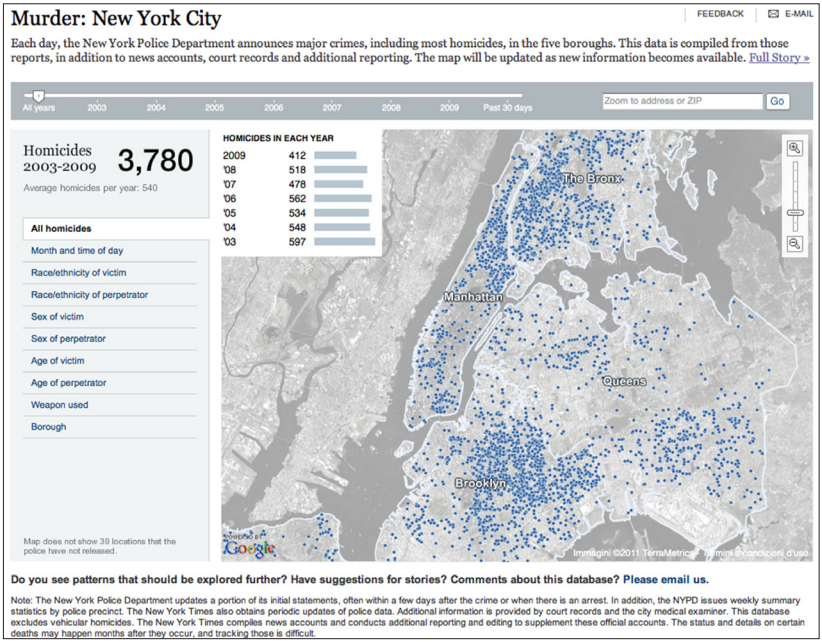
▶▶ fig 20a-b | Murder: New York City. Homicides 2003-2009
Immagine tratta dal sito http://projects.nytimes.com/crime/homicides/map

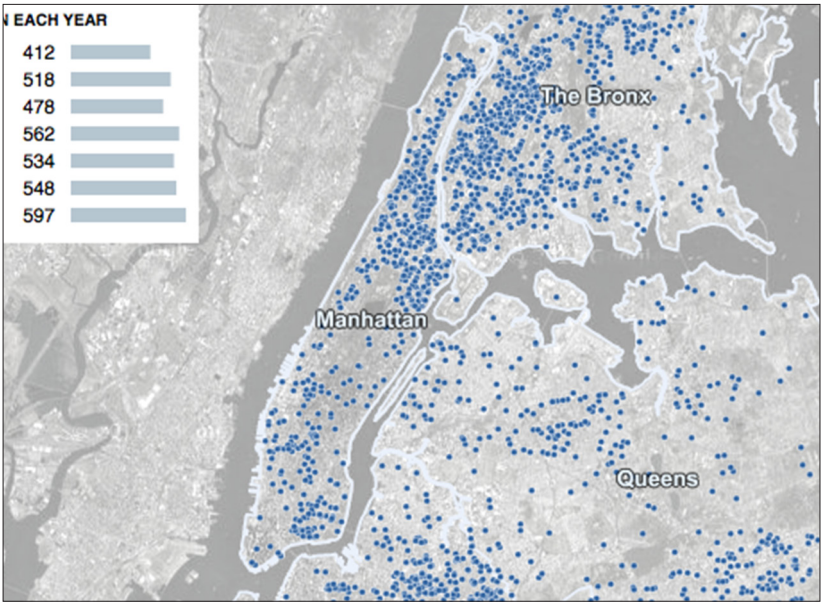
▶▶▶ fig 21a-b | The Oxford Crime Map
Immagine tratta dal sito http://www.bbc.co.uk/truthaboutcrime/crimemap/



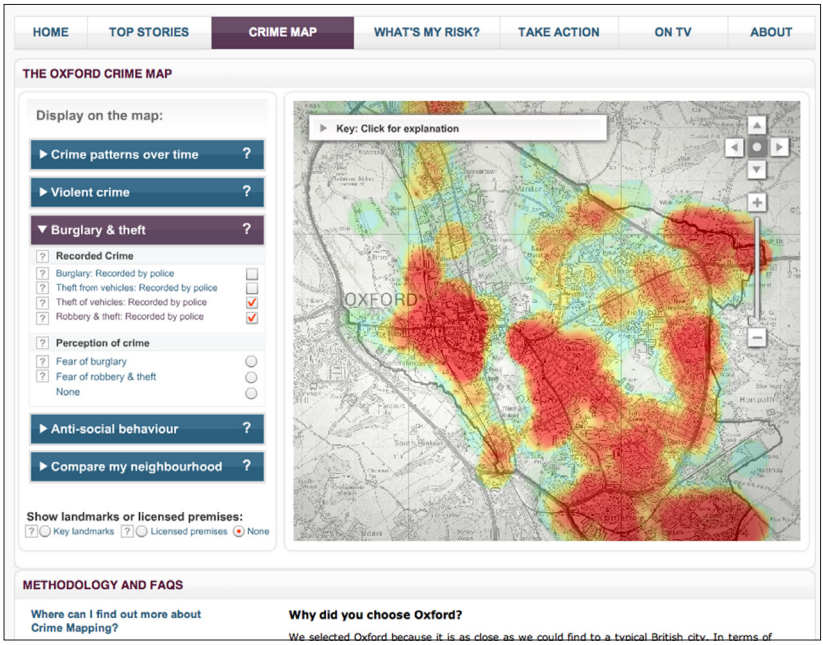
3. Visualizzare le minacce informatiche: l'attuale stato dell'arte

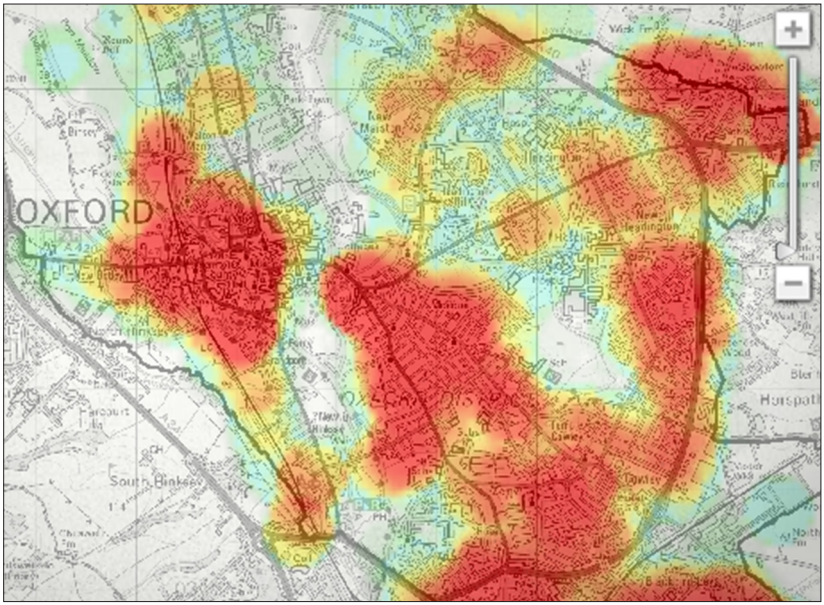

Home
Map
Crime Reports
Police Beats
Alerts
API
About
Feedback
CRIME TYPE
Show All | Hide All
AA Aggravated Assault
Mu Murder
Ro Robbery
SA Simple Assault
DP Disturbing the Peace
Na Narcotics
Al Alcohol
Pr Prostitution
Th Theft
VT Vehicle Theft
Va Vandalism
Bu Burglary
Ar Arson
PIEDMONT
Dingee Reservoir
Upper Broadway
Lakeside
Laney College
Highland Park





3. Visualizzare le minacce informatiche: l'attuale stato dell'arte


HOME
TOP STORIES
CRIME MAP
WHAT'S MY RISK?
TAKE ACTION
ON TV
ABOUT
THE OXFORD CRIME MAP
Display on the map:
Crime patterns over time
Violent crime
Burglary & theft
Recorded Crime
Burglary: Recorded by police
Theft from vehicles: Recorded by police
Theft of vehicles: Recorded by police
Robbery & theft: Recorded by police
Perception of crime
Fear of burglary
Fear of robbery & theft
None
Anti-social behaviour
Compare my neighbourhood
Show landmarks or licensed premises:
Key landmarks
Licensed premises
None
Key: Click for explanation
OXFORD
METHODOLOGY AND FAQS
Where can I find out more about Crime Mapping?
Why did you choose Oxford?
We selected Oxford because it is as close as we could find to a typical British city. In terms of


OXFORD

4 La visualizzazione come strumento per la ricerca e l'analisi

"Modern data graphics can do much more than simply substitute for small statistical tables. At their best, graphics are instruments for reasoning about quantitative information. Often the most effective way to describe, explore, and summarize a set of numbers -even a very large set- is to look at pictures of those numbers."

Edward R. Tufte,
The Visual Display of Quantitative Information (second edition)

Nella vita di tutti i giorni siamo sempre più circondati e sommersi da vari tipologie di visualizzazioni: dalle più popolari infografiche e mappe presenti su riviste, giornali e siti web a quelle più specifiche e complesse tipiche degli ambienti lavorativi che servono ad illustrare e spiegare workflow, processi, strutture ed organizzazioni complesse. In questa sovrabbondanza di informazione visiva gli strumenti di visualizzazione hanno man mano preso sempre più piede in molteplici discipline, diffondendosi nei processi decisionali, nelle fasi di pianificazione, nelle presentazioni aziendali. Se da un lato questo può sicuramente essere considerato come un aspetto positivo verso un'attenzione alla comunicazione, dall'altra parte il diffondersi di questa pratica ha portato alla creazione di un vasto insieme di visualizzazioni realizzate da non esperti del settore. Questo ha fatto sì che si creasse un mondo frammentato, ricco di svariate tipologie di prodotti, dove, nonostante la profonda conoscenza di uno specifico settore porti a conoscere molto bene quali sono i principali bisogni che vogliono essere visualizzati, lo spazio dedicato alla ricerca sui metodi di visualizzazione si è ridotto all'utilizzo di forme e soluzioni già conosciute, di volta in volta riadattate secondo le esigenze, e che puntano più a creare un report visivo dei dati a disposizione rispetto che ad un vero e proprio strumento di ricerca ed analisi visiva.

Per far fronte a queste problematiche diverse discipline, come ad esempio l'Information Visualization, e autori, vedi Edward R. Tufte, hanno sviluppato

linee di guida e di pensiero sul come realizzare al meglio una visualizzazione, prendendo anche qui svariate strade e tipologie di approccio (più o meno funzionale ed analitico, o più o meno comunicativo-artistico come ad esempio nella information-aesthetics, e così via). Si tratta però sempre di discipline isolate, utili nelle specifiche visualizzazioni, ma che non garantiscono un approccio di tipo progettuale. In quest'ottica la figura del designer della comunicazione può inserirsi in maniera interessante, con un occhio sempre rivolto al contesto in cui la visualizzazione si inserisce come strumento di visualizzazione, con un'attenzione verso i processi e i metodi di interazione tra l'utente ed il prodotto, e con tutte le altre sue specifiche competenze, giocando un ruolo importante.

"This may be especially true if -the challenges of the modern world require integrative problem solving and, at a more comprehensive level, holistic thought and transdisciplinary schema promote unity of knowledge[1]- and given that -without integrative disciplines of understanding, communication, and action, there is little hope of sensibly extending knowledge beyond the library or laboratory in order to serve the purpose of enriching human life[2]-. It is exactly in this way of thinking that design -by nature an interdisciplinary, integrative discipline[3]- and -the ability of designers to discover new relationships among signs, things, actions, and thoughts [...][3]- can play a big role.

Working on visualizations from a communication design perspective requires a different approach, one able to consider them in the wider domain of communication strategy. This is the reason why we need a model able to cope with the many nuances of visualizations, a model capable of taking into account the context in which visualizations act as communication tools.

In such a design perspective we must start to refer to visualizations as means to achieve purposes. This does not refer strictly to the idea of representing of high-dimensional sets of data: -the ability to visualize complex information does not refer solely to the communication of quantitative information but it also deals with the visual

[1] J.T. Klein. Crossing boundaries: Knowledge, disciplinarities, and interdisciplinarities. Charlottsville: The University of Virginia Press. 1996

[2] R. Buchanan "Wicked Problems in Design Thinking". Design Issues, 8:2, 5-21, Cambridge, MA: The MIT Press. 1992

[3] K. Friedman. "Theory construction in design research: criteria: approaches, and methods", Design Studies 24:6, November. 2003

narration of values and qualitative data[4]-. So the aims of a visualization can include indeed the capability to make sense of context, communicate impressions, telling stories. We need to change our perspective so that visualizations are not merely defined by the technology they involve, but rather by the relation with the aim and context they are designed for and the recipient they want to reach." [5]

L'obiettivo di una buona visualizzazione è quindi la capacità di dare senso ai dati utilizzati per crearla, in riferimento al contesto comunicativo in cui si inserisce, ed eventualmente essendo capace di trasmettere impressioni, concetti e contenuti altrimenti incomunicabili attraverso i soli dati. Vista in quest'ottica, a prescindere dalle tecniche utilizzate per realizzarla, può divenire quindi ottimo strumento per la ricerca e l'analisi, con l'obiettivo di generare nuovi spunti e conoscenze sia in coloro i quali la progettano, sia, e soprattutto, in coloro i quali ne diverranno gli utenti finali.

4.1 Il modello data-information-knowledge, la visualizzazione come processo

Presentato nel paper "From Data to Knowledge - Visualizations as transformation processes within the Data-Information-Knowledge continuum (2010)" [5] questo modello posa i suoi fondamenti partendo dal presupposto che le visualizzazioni, come quelle prese in analisi fino a questo momento, lavorino sì partendo da una base di dati, ma non solo: le visualizzazioni lavorano, e si rapportano a chi le sta osservando, anche tramite l'informazione e la conoscenza.

Per gli operatori del settore non sarà difficile essersi già imbattuti in termini come Data Visualization, Information Visualization o Knowledge Visualization, di volta in volta utilizzati come nome di una disciplina in cui inserire un determinato

[4] G. Scagnetti, D. Ricci, G. Baule, P. Ciuccarelli. "Reshaping Communication Design Tools" Complex Systems Structural Features For Design Tools, in IASDR07: "International Associations of Societies of Design Research" The Hong Kong Polytechnic University, Hong Kong, 20 (digital). 2007
[5] L. Masud, F. Valsecchi, P. Ciuccarelli, D. Ricci, and G. Caviglia. From Data to knowledge - visualizations as transformation processes within the data-information-knowledge continuum. Information Visualization, Intl. Conference on, 0:445-449, 2010

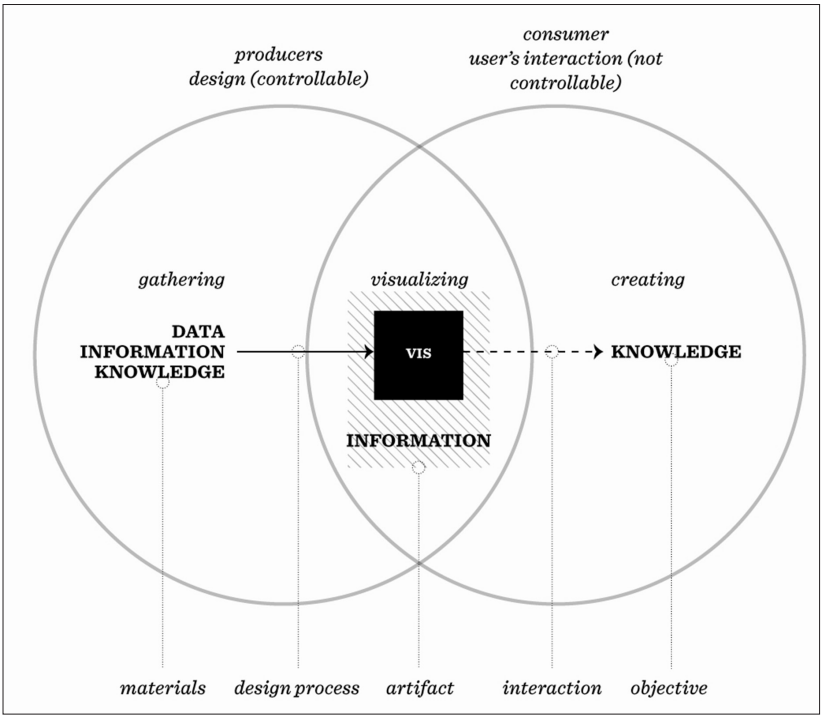
progetto di visualizzazione. Il modello data-information-knowledge presenta un processo, in cui a partire da questi tre elementi, attraverso la visualizzazione, si va a creare nuova conoscenza.

"As Bellinger, Castro and Mills [6] refer, the continuum starts with raw data, -it simply exists and has no significance beyond its existence (in and of itself). It can exist in any form, usable or not-, proceeds with information that occurs when data has been given meaning by way of relational connections. Also in this case, -This 'meaning' can be useful, but does not have to be-. Finally knowledge is the appropriate collection of information, -such that its intent to be useful-. For example data may be the different temperatures measured by a thermometer. By relating these different temperatures it is possible to understand if the temperature is rising or decreasing, giving the data a meaning and thus transforming it in information (such that it may be useful). Lastly, the fact that it is known that the temperature is decreasing may trigger on action: e.g. if the temperature was measured at home then the heating might be turned on." [5]

Il processo inizia quindi dai dati grezzi, che presi singolarmente esistono di per sé ma sono privi di significato. Questi si possono trovare in qualsiasi forma, più o meno adatta ad essere utilizzata, e quando vengono messi in relazione tra loro, attraverso delle connessioni, iniziano a generare informazione. L'informazione così generata può essere di due tipi: utile o non utile. La conoscenza deriva dalla raccolta di tutte le informazioni utili che è stato possibile ricavare dai dati iniziali. Vista sotto questo punto di vista una visualizzazione non si limita ad essere un prodotto di rappresentazione di un dato set di dati, bensì diviene processo di trasformazione degli stessi per generare nuova informazione e conoscenza. Come rappresentato in fig. 22, le visualizzazioni possono infatti raccogliere dati, informazione o conoscenza (i materiali di partenza), rappresentarli come informazione attraverso una visualizzazione (un artefatto), ed eventualmente creare nuova conoscenza (obiettivo). Il processo si compone di due parti: il progettista che genera la rappresentazione ed il consumatore. Il primo è colui che agisce attivamente nella progettazione della visualizzazione: visualizzare vuol dire selezionare ed ordinare i dati e le informazioni che si hanno, al fine di decidere come

[6] G. Bellinger, D. Casto and A. Mills. "Data, information, knowledge, and wisdom". http://www.system-thinking.org/dikw/dikw.htm. 2004

e cosa mostrare del materiale di partenza. La seconda parte del processo è invece interamente nelle mani del consumatore, che interagendo con la visualizzazione acquisisce conoscenza.



▲ **fig 22 | Visualizations as processes within the DIK continuum.1**
Immagine tratta dal paper From Data to knowledge - visualizations as transformation processes within the data-information-knowledge continuum.

Scopo ultimo del processo è quindi il tentativo di creare nuova conoscenza per il consumatore, e per fare ciò si rende necessario un approccio che abbia sempre sott'occhio gli scopi e gli obiettivi ultimi della visualizzazione che si sta andando a progettare. Poiché diverse tipologie di visualizzazioni sono caratterizzate da differenti funzionalità, alcune visualizzazioni possono essere più adatte di altre a rappresentare determinate condizioni ed ogni processo di visualizzazione genera differenti tipologie di conoscenza, il modello data-information-knowledge porta con se anche un ragionamento su quali visualizzazioni possono essere utili per rappresentare cosa.

Basandosi anche sugli scritti di autori come G. Schraw [7] e G. Judelman [8] viene fatta una distinzione tra tre differenti tipologie di conoscenza: conoscenza dichiarativa, conoscenza procedurale e conoscenza condizionale.

La conoscenza dichiarativa, conoscere il quale o conoscere il che cosa, include tutte quelle visualizzazioni che partendo da un set di dati, più o meno astratti, li convertono in informazione, con l'obiettivo di permettere all'utente di acquisire nuova conoscenza e costruire ipotesi a partire dai dati.

La conoscenza procedurale, conoscere il come, è riferita a quelle visualizzazioni che, come ad esempio le infografiche sviluppate per riviste e quotidiani, non si limitano a prendere dei dati e visualizzarli, ma che attraverso relazioni e connessioni tra questi divengono in grado di "raccontare una storia" ed in questo modo comunicare un certo tipo di informazione. Questa informazione può essere utilizzata da chi la osserva sia per comprendere un fenomeno che per comprendere come fare qualcosa, ad esempio nel caso in cui la visualizzazione venga usata per generale dei manuali di istruzioni. Le visualizzazioni inserite in questo gruppo quindi non si limitano soltanto a fornire all'utente una conoscenza di tipo dichiarativo, ma producono conoscenza procedurale.

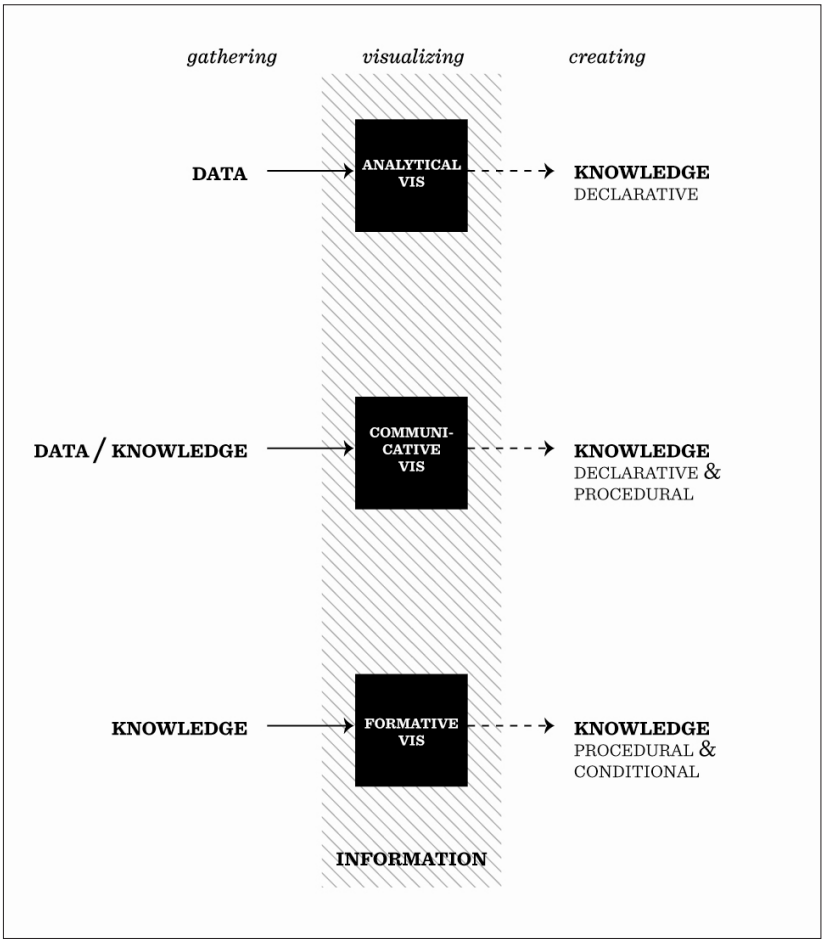
[7] G. Schraw. "Promoting general metacognitive awareness". Instructional Science 26, no. 1: 113-125. 1998
[8] G. Judelman. Knowledge Visualization. Problems and Principles for Mapping the Knowledge Space. M. Sc. Thesis. 2004

Infine la conoscenza condizionale, conoscere quando e perché, è quel processo che non basa più il suo punto di partenza sui dati, e, poiché scopo ultimo della Knowledge Visualization non è quello di dare senso a qualsivoglia ampio e complesso set di dati, bensì generare conoscenza condivisa, non si limita soltanto a comunicare come fare le cose (conoscenza procedurale), bensì ha come obiettivo quello di trasferire conoscenza sul quando e perché utilizzare la conoscenza stessa.

Queste tre categorie di conoscenza possono essere riportate in tre categorie di visualizzazione: visualizzazioni analitiche, visualizzazioni comunicative e visualizzazioni formative.

Visualizzazioni analitiche Questa tipologia di visualizzazioni, caratterizzate da un approccio ti tipo funzionale, fa largo uso di tecniche reversibili in maniera tale che colui che osserva la visualizzazione sia in ogni momento in grado di dedurre da quest'ultima i dati con i quali è stata generata. Generalmente vengono utilizzate con scopi prettamente tecnici e, principalmente, da coloro che ogni giorno hanno a che fare, per lavoro, con un qualche tipo ti rappresentazione visiva: esperti ed operatori di vari settori quali possono essere ad esempio analisti, economi, statistici, personale medico, e così via.

Visualizzazioni comunicative Come per la conoscenza procedurale, in questo caso le visualizzazioni non vengono usate semplicemente per visualizzare dati ma servono a comunicare e raccontare le relazioni tra i dati stessi e i significati che queste relazioni portano con sé: chi progetta la visualizzazione agisce come intermediario nel processo di comunicazione, raccontando dei risultati, comunicando un fenomeno. Per questo motivo queste visualizzazioni sono capaci di trasmettere consapevolezza, conoscenza sociale, spingere alla riflessione ed all'approfondimento. Inoltre, essendo distribuite tramite i mass media, ed essendo quindi potenzialmente accessibili a tutti, compresa tutta quella fascia di utenza non in possesso delle conoscenze pregresse adatte ad una loro completa comprensione, questa tipologia di visualizzazioni spesso utilizzano linguaggi pittogrammatici, metafore ed illustrazioni. In questo contesto, non essendo il recupero di dati precisi uno degli obiettivi primari, un approccio basato su un rapporto



▲ **fig 23 | Visualizations as processes within the DIK continuum.2**
Immagine tratta dal paper From Data to knowledge - visualizations as transformation processes within the data-information-knowledge continuum.

di tipo data-ink [9] potrebbe risultare inefficace o addirittura controproducente.

Visualizzazioni formative Questo terzo gruppo di visualizzazioni viene profondamente influenzato dal contesto in cui le visualizzazioni stesse vengono usate. In questo gruppo rientrano difatti tutte quelle visualizzazioni che condividono tecniche di comunicazione anche con obiettivi o contesti di utilizzo molto diversi tra loro. Sono spesso utilizzate come supporto nel trasferimento di conoscenza all'interno dei gruppi di lavoro. Possono essere rappresentazioni di processi, flussi di lavoro, strutture amministrative, e sono in grado di istruire i singoli utenti sul loro ruolo all'interno di quel contesto lavorativo. L'elemento principale che caratterizza questo tipo di visualizzazioni è l'azione: sono visualizzazioni create per persone che ricoprono ruoli attivi all'interno di un'organizzazione od un gruppo di lavoro in maniera tale che questi ultimi siano messi in grado di conoscere come, quando e perché agire in un determinato contesto.

Per concludere, visualizzare può essere quindi a tutti gli effetti considerato come un processo in cui a partire da un dato set di dati, informazioni e conoscenze, attraverso diversi percorsi, si strutturano, legano, relazionano ed infine visualizzano i materiali di partenza con lo scopo ultimo di generare nuova conoscenza.

4.2 Non solo visualizzare dati, ma raccontare fenomeni

Come già anticipato nel capitolo sui benefici legati alla visualizzazione la rappresentazione visiva di un set di dati permette di comunicare, e raccontare, a chi la osserva una grande quantità di informazioni, molto maggiore di quella che si potrebbe comunicare in uno stesso lasso di tempo, attraverso un semplice elenco o report testuale. Difatti, mentre per il nostro cervello è molto facile ed intuitivo comprendere ed assimilare figure ed immagini, risulta più lungo e energicamente dispendioso processare un testo. Tutto questo, combinato alla capacità che le immagini hanno di contenere grandi quantità di informazioni in uno spazio relativamente compresso, grazie alle grandi quantità di attributi

[9] Edward R. Tufte. "The Visual Display of Quantitative Information". Second Edition. Graphic Press, Cheshire, Connecticut. Sixth printing, august 2009

e variabili visive che possono essere messe in gioco, rende la visualizzazione uno strumento di grande utilità ed importanza nella trasmissione della conoscenza. Inoltre come abbiamo visto nel capitolo precedente quando abbiamo parlato di *Visualizzazioni Comunicative*, alcune tipologie di visualizzazioni sono particolarmente utili nei processi di comunicazione, comunicando e raccontando un fenomeno, altrimenti difficile, o addirittura in alcuni casi impossibile, da cogliere avendo come unico riscontro il mezzo testuale. Visualizzare vuol quindi anche dire dare la possibilità a chi osserva di avere tra le mani uno strumento in grado di metterlo in condizione di comprendere meglio un dato fenomeno, fare nuove ipotesi e relazioni tra i dati che sta osservando, e coglierne a pieno tutto il carico conoscitivo. Inoltre osservare vuol dire anche aiutare a fissare maggiormente nella memoria dati e fenomeni visualizzati.

Nel corso della storia numerose visualizzazioni sono state utilizzate a questo scopo, riporto qui di seguito alcuni esempi tra i più classici e conosciuti, ma forse proprio per questo tra i più significativi e adatti ad essere utilizzati in maniera esemplificativa.

L'epidemia di colera nella Londra del 1854

Nel suo libro *On the Mode of Communication of Cholera*, John Snow descrive la pesante epidemia che colpì la città di Londra nel 1854:

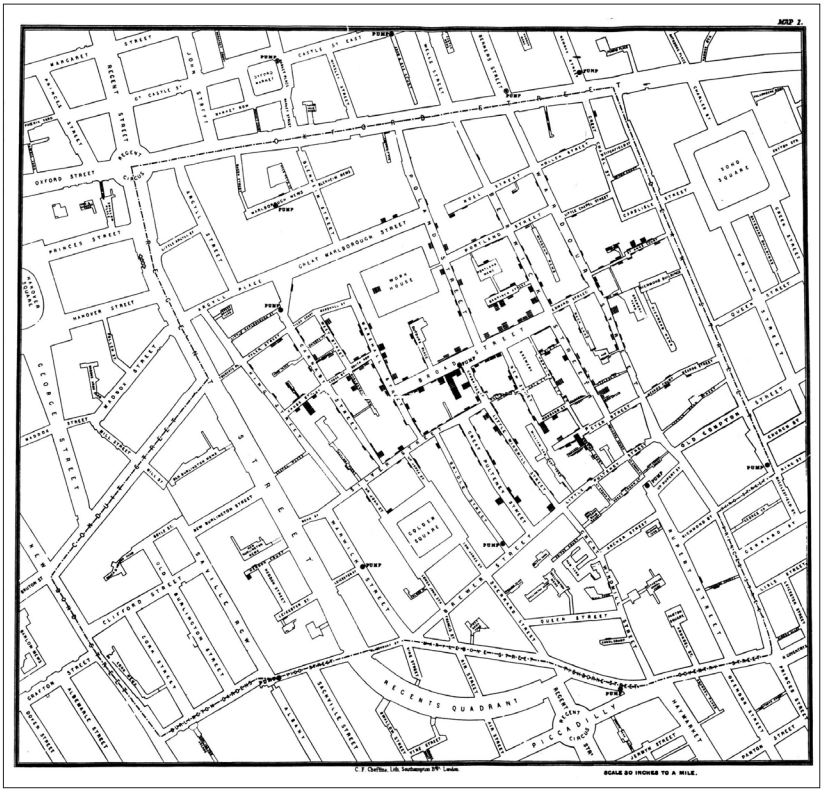
"The most terrible outbreak of cholera which ever occurred in this kingdom, is probably that which took place in Broad Street, Golden Square, and adjoining streets, a few weeks ago. Within two hundred and fifty yards of the spot where Cambridge Street joins Broad Street, there were upwards of five hundred fatal attacks of cholera in ten days. The mortality in this limited area probably equals any that was ever caused in this country, even by the plague; and it was much more sudden, as the greater number of cases terminated in a few hours. The mortality would undoubtedly have been much greater had it not been for the flight of the population. Persons in furnished lodgings left first, the other lodgers went away, leaving their furniture to be sent for. [...] Many houses were closed altogether owing to the death of the proprietors; and, in a great number of instances, the tradesmen who remained had sent away their families; so that in less than six days from the commencement of the outbreak, the most afflicted streets were deserted by

more than three-quarters of their inhabitans." [10]

L'epidemia eruppe nell'area della Broad Street di Londra sul finire del mese di agosto del 1854. John Snow, rifacendosi ad altre sue ricerche già fatte su precedenti epidemie, iniziò a sospettare che l'alta mortalità dell'area potesse essere in qualche modo connessa con l'approvvigionamento idrico. Dopo varie ricerche e tentativi, presa una mappa e segnate su di essa i punti in cui erano state registrate le morti e i punti in cui erano presenti le pompe per l'acqua, risultò evidente che la più alta concentrazione di morti si era verificata attorno ad una pompa situata proprio lungo la Broad Street. Rimossa quella l'epidemia iniziò rapidamente a decrescere. Si tratta di un esempio importante in quanto bisogna pensare che fino ad allora le credenze attribuivano la diffusione del colera ad una trasmissione aerea di un male che proveniva dal terreno dove secoli prima vi era stata la peste. Inoltre i dati raccolti sulla mortalità e la diffusione del fenomeno venivano raccolti in ordine di data di morte, e riportavano i nomi delle vittime e le circostanze in cui era avvenuto il decesso. Questo metodo di raccolta dei dati forniva un eccellente narrazione cronologica degli avvenimenti ma non aiutava in nessuna maniera la ricerca sulle cause del fenomeno. La grande intuizione di Snow, quella di prendere una mappa e fisicamente segnare su di essa tutti i punti in cui le morti erano avvenute, ed i punti in cui erano posizionate le pompe per l'approvvigionamento idrico, fu di grande importanza nel mostrare, raccontare, alle autorità come i due fenomeni potessero essere tra loro collegati, creandone una chiara relazione visiva, e nel convicerle a tentare una nuova linea di azione, che portò ad una rapida risoluzione del problema.

▶ fig 24 | John Snow, la mappa del colera
Immagine tratta dagli scritti di John Snow, On the Mode of Communication of Cholera (Londra, 1855)

[10] John Snow, "On the Mode of Communication of Cholera", p.38. Londra. 1855



L'armata francese nella campagna di Russia, 1812-1813

Un altro chiaro esempio di come la visualizzazione risulti utile per la sua capacità, e la sua forza, nel raccontare un fenomeno è la mappa creata nel 1869 da Joseph Minard che rappresenta la campagna francese in Russia del 1812-1813.

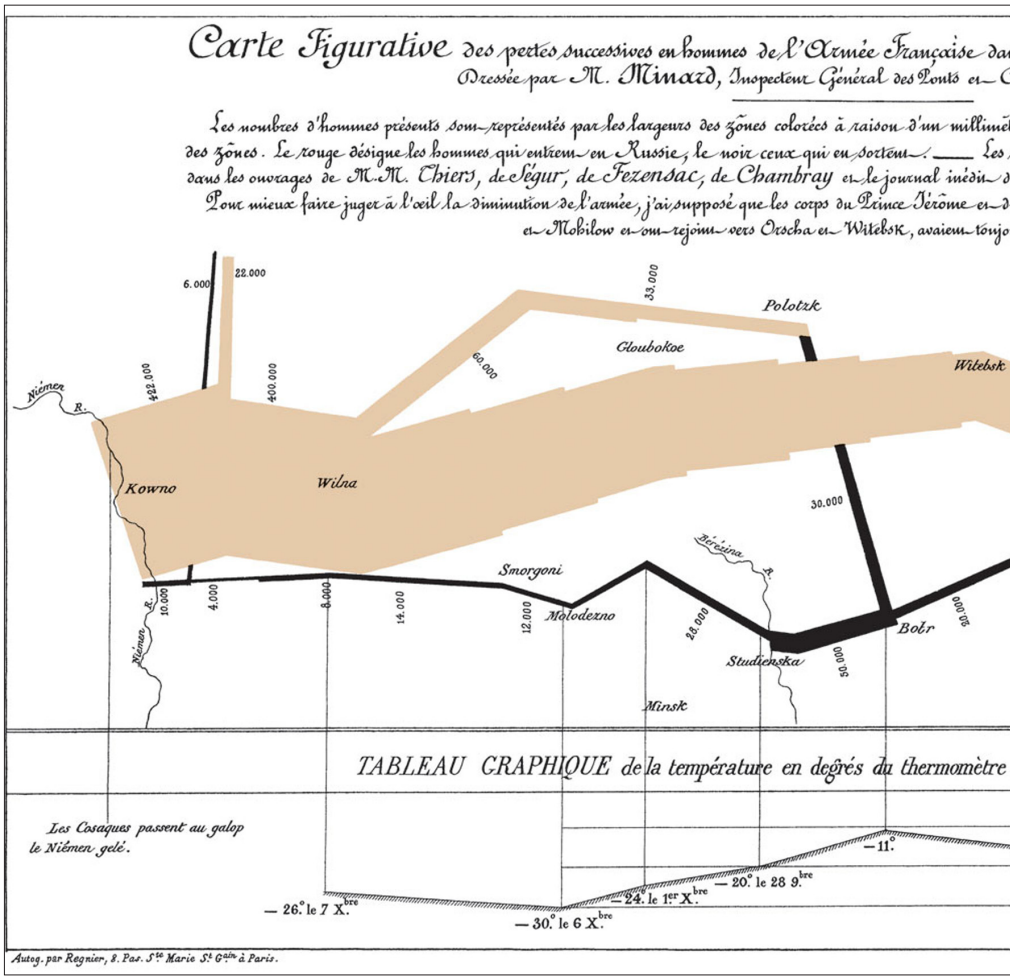
Joseph Minard, ingegnere francese, nel 1869, rifacendosi ai dati conosciuti circa la disastrosa campagna di Russia dell'esercito napoleonico, creò una visualizzazione della stessa, per raccontarne visivamente l'andamento. La mappa che ne scaturì venne descritta da E. J. Marey come un qualcosa *"seeming to defy the pen of the*

[11] E. J. Marey, "La méthode graphique". Paris. 1855
[12] Edward R. Tufte. "The Visual Display of Quantitative Information". Second Edition. Graphic Press, Cheshire, Connecticut. Sixth printing, august 2009

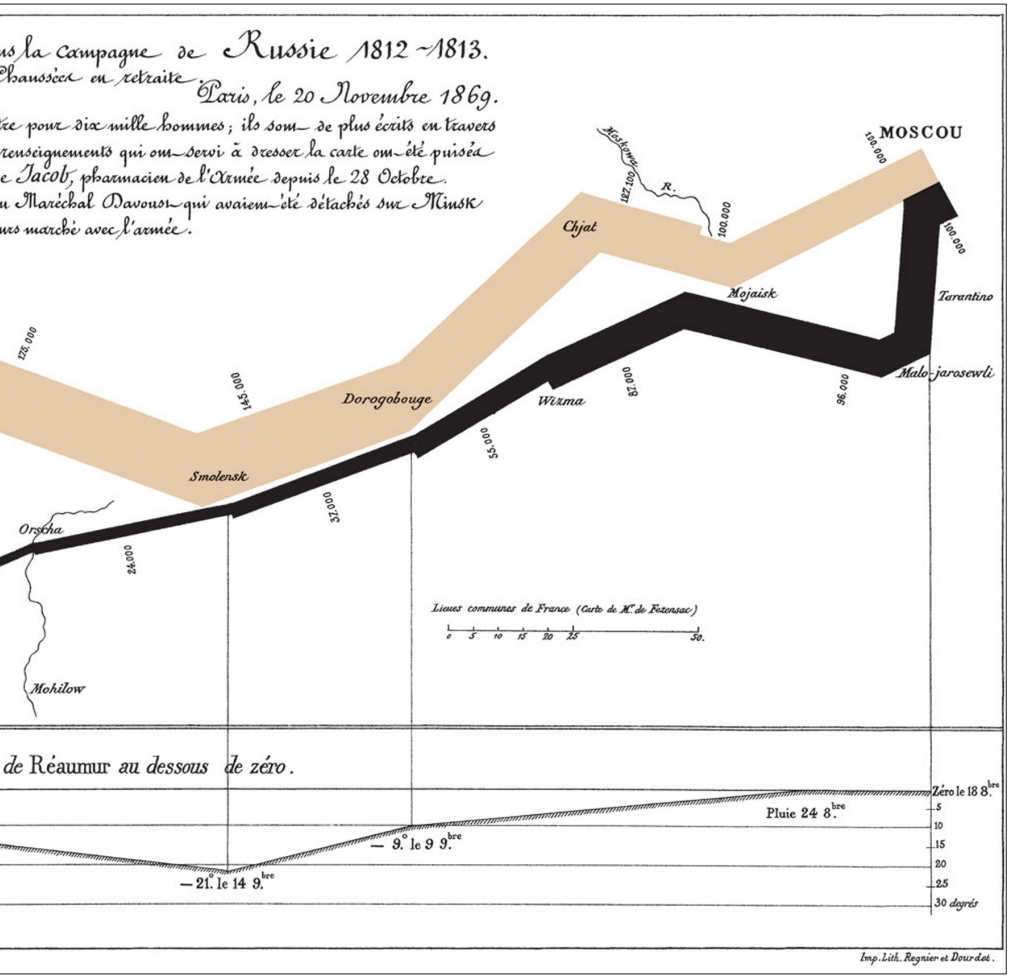
historian by its brutal eloquence" [11], ed in seguito a detta di Edward R. Tufte *"it may well be the best statistical graphic ever drawn."* [12].
Letta da sinistra verso destra la mappa inizia lungo le rive del fiume Niemen, al confine tra Polonia e Russia. La linea marrone disegna la strada fatta dall'esercito napoleonico nel suo percorso verso mosca, posizionata in alto a destra. Lo spessore della linea mostra la dimensione dell'armata napoleonica, composta, nel giugno del 1812, all'inizio dell'invasione, da 422.000 unità. Lungo il percorso lo spessore di questa linea mostra il numero, la dimensione dell'armata napoleonica, registrata tappa per tappa, che gradualmente diviene sempre più sottile. All'arrivo a Mosca, nel settembre del 1812, una città già saccheggiata e deserta, l'armata dell'esercito napoleonico ammontava a 100.000 unità, tre quarti degli uomini che lo componevano erano già morti lungo il percorso. Da destra verso sinistra, con una linea nera, viene rappresentata la conseguente ritirata ed il tentativo dell'esercito di ritornare verso la patria. Questa linea è anche connessa con la parte inferiore della visualizzazione dove una linea indica date e temperature, sempre sotto lo zero. Sempre ben visualizzato dalla mappa sono i disastrosi attraversamenti di fiumi, come ad esempio il Berezina sulla via del ritorno nella seconda metà di novembre, dove presumibilmente quasi la metà degli uomini rimasti morirono nel tentativo di attraversarlo. Quando l'esercito riuscì a rientrare in Polonia contava di sole 10.000 unità.

La visualizzazione di Minard racconta con una forza visiva irraggiungibile da un elenco testuale di dati la realtà di questa campagna, con una carica emotiva che difficilmente si potrebbe pensare di attribuire ad un grafico.

▶ fig 25 | L'armata francese nella campagna di Russia, Joseph Minard
Immagine tratta dal Tableaux Graphiques et Cartes Figuratives de M. Minard, 1845-1869, Bibliothèque de l'École Nationale des Ponts et Chausséès, Paris.



4. La visualizzazione come strumento per la ricerca e l'analisi


...s la campagne de Russie 1812-1813.
...haussée en retraite
Paris, le 20 Novembre 1869.
...re pour dix mille hommes ; ils sont de plus écrits en travers
...renseignements qui ont servi à dresser la carte ont été puisés
...e Jacob, pharmacien de l'Armée depuis le 28 Octobre.
...u Maréchal Davoust qui avaient été détachés sur Minsk
...ns marché avec l'armée.
MOSCOU
100.000
Moskowa
R.
127.100
100.000
Chjat
Mojaisk
Tarantino
100.000
Malo-jarosewli
96.000
82.000
Wizma
54.000
Dorogobouge
145.000
125.000
Smolensk
37.000
Orscha
24.000
Mohilow
Lieues communes de France (Carte de M.r de Fezensac)
0 5 10 15 20 25 50.
...de Réaumur au dessous de zéro.
Zéro le 18 8.bre
Pluie 24 8.bre
5
10
15
20
25
30 degrés
— 9.° le 9 9.bre
— 21. le 14 9.bre
Imp. Lith. Regnier et Dourdet.

4.3 Oltre il concetto di dashboard

Il concetto di dashboard nasce nel 1980 sotto il nome di EISs, Executive Information Systems, quando a scopo lavorativo, in uffici e aziende, iniziano a circolare alcuni programmi che hanno come obiettivo quello di monitorare diverse misure finanziare attraverso l'utilizzo di un'unica schermata, che racchiuda tutto ciò che può essere utile tenere in vista, e che sia semplice ed intuitiva in maniera tale da poter essere utilizzata e compresa anche dagli operatori meno esperti. In quegli anni però la tecnologia non era ancora sufficiente a fornire i mezzi adatti a supportare un simile sistema, e bisognerà aspettare fino al decennio successivo per iniziare a vederne le prime significative applicazioni. Entrato ormai nella terminologia comune quando si parla di dashboard si ha già un'idea di massima di a che cosa ci si stia riferendo, e molti di noi associano questo termine ad applicazioni che, utilizzando interfacce ricche di tabelle, grafici, indicatori visivi, ci ritornano all'interno di un unica schermata una grande varietà di informazioni e dati, non sempre legate tra loro. Tipicamente una dashboard si compone di un'unica schermata suddivisa in sottosezioni, dove all'interno di ogni sezione viene monitorato un determinato dato. Questo però non è sufficiente ad identificare e stabilire cosa rientri in questa tipologia di strumenti, e per questo si rende necessario, volendo approfondire maggiormente l'argomento, riportare la definizione di dashboard proposta da Stephen Few:

"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance." [13]

E ancora:

"Just as the dashboard of a car provides critical information needed to operate the vehicle at a glance, a BI dashboard serves a similar purpose, whether you're using it to make strategic decisions for a huge corporation, run the daily operations of a team, or performs tasks that involve no one but yourself. The means is a single-screen display, and the purpose is to efficiently monitor the information needed to achieve one's objectives." [14]

[13] Stephen Few. "Dashboard Confusion". Intelligent Enterprise (magazine). 20 Marzo 2004
[14] Stephen Few. "Information Dashboard Design. The Effective Visual Communication of Data". O'Reilly. 2006

Quindi una dashboard si compone come un *visual display*, dove l'informazione viene rappresentata visivamente, solitamente come una composizione di testo e grafica, con una particolare enfasi verso quest'ultima. La scelta di enfatizzare la parte grafica è dettata dall'efficienza ed immediatezza che caratterizza la comunicazione per immagini e la percezione visiva. Essa mostra *"the most important information needed to achieve one or more objectives"*: il raggiungimento di un singolo obiettivo si compone sempre di più parti, a volte apparentemente slegate tra loro, ma tutte utili alla comprensione del fenomeno nel suo insieme; non c'è un tipo specifico di informazione, ma tutte le informazioni messe assieme collaborano al raggiungimento dello scopo. Si compone affinché possa essere visualizzata *"on a single computer screen"*, per garantire a chi la osserva l'accessibilità immediata a tutta l'informazione contenuta, senza dover cambiare pagina o scrollare la stessa, ed in maniera tale che essa possa sempre essere *"monitored at a glance"*.

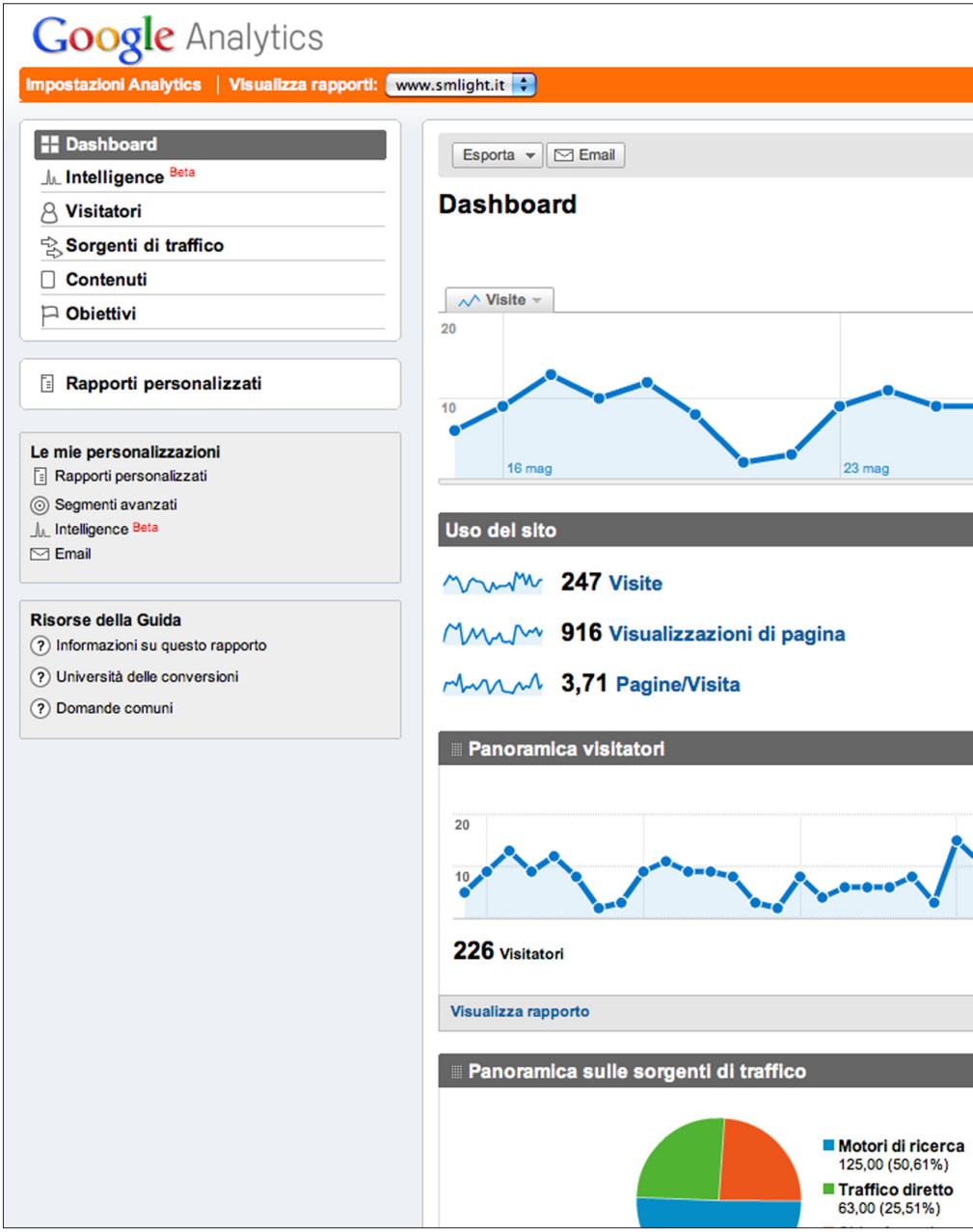
In aggiunta a questo, nel libro *"Information Dashboard Design"*, Stephen Few propone anche una serie di parametri e caratteristiche per la realizzazione di una buona dashboard. In particolare evidenzia tredici errori che spesso si possono riscontrare nelle dashboard presenti sul mercato:

Superare i confini di un unica schermata Se un layout non è progettato per essere visualizzato nella sua interezza all'interno di un'unica schermata, ma necessita di essere scrollato per visualizzare tutte le zone della dashboard, costringerà l'utente a dover vedere soltanto una parte per volta dell'intera dashboard, celandogli di volta in volta informazioni e limitandone la possibilità di avere una visione completa d'insieme.

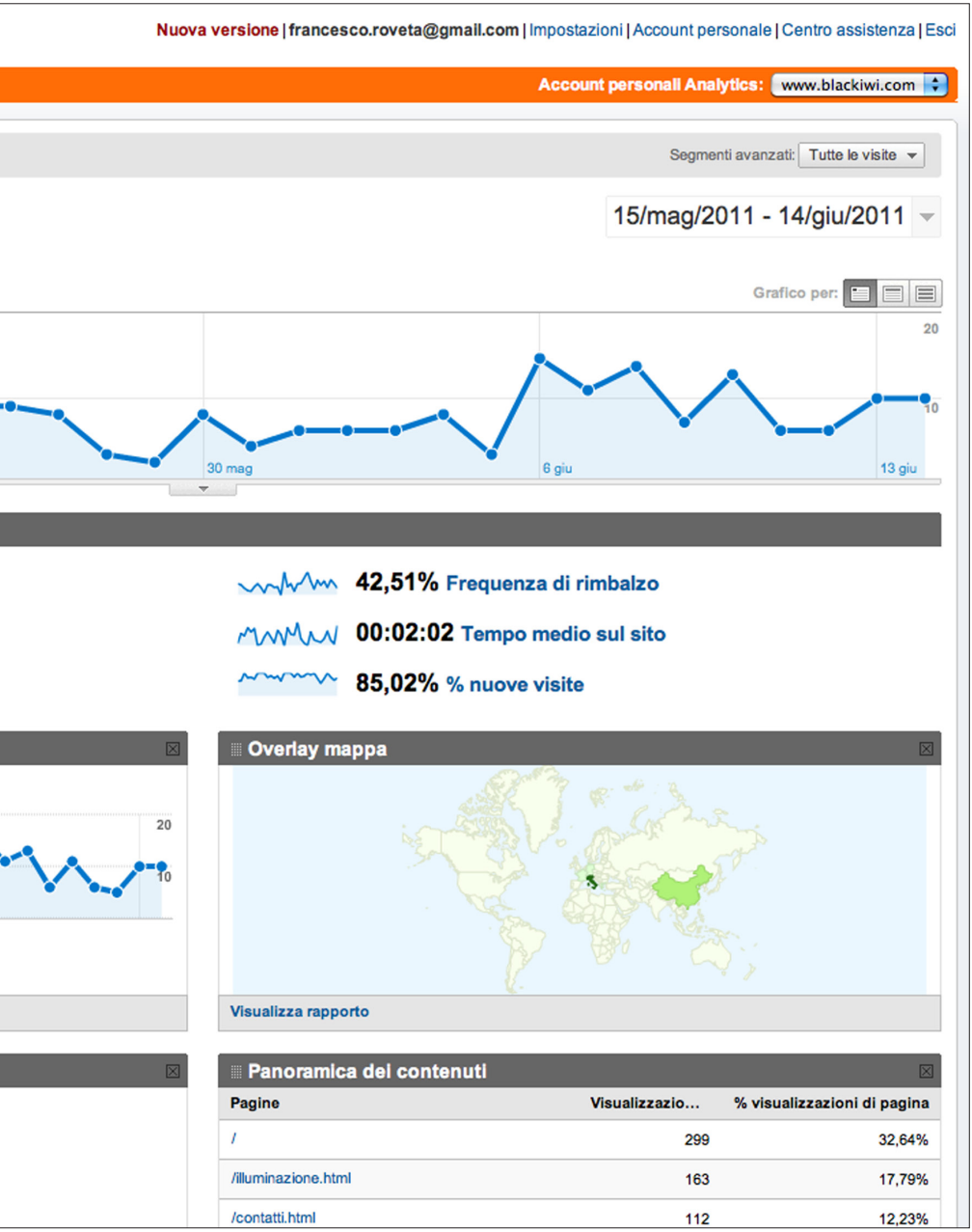
Fornire un contesto inadeguato per i dati Il singolo dato di per se non fornisce informazione. L'informazione si genera dai dati quando questi vengono messi in

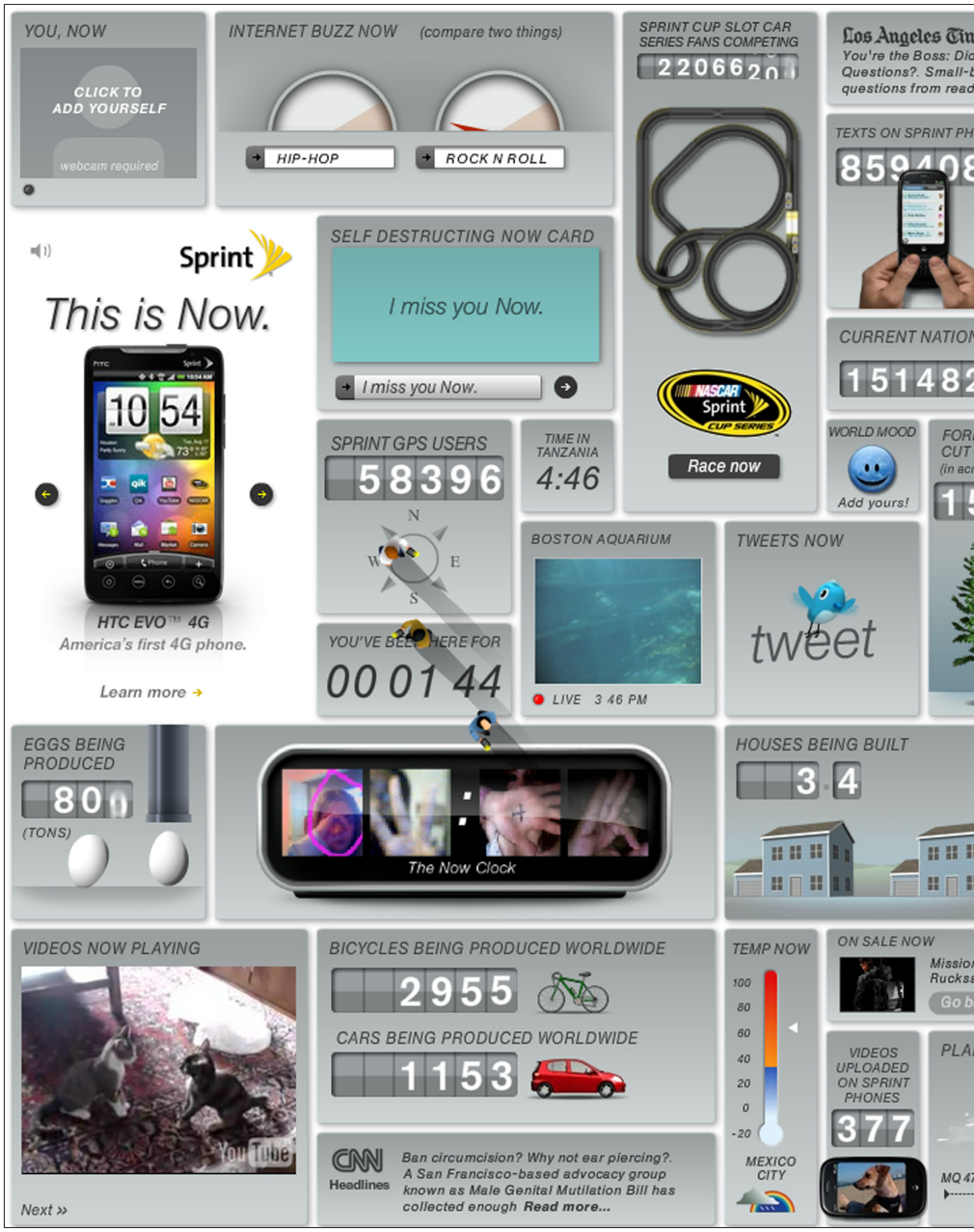
▶ **fig 26 | Google Analytics Dashboard**
Immagine tratta da un'area privata del sito http://www.google.com/analytics/

▶▶ **fig 27 | Now Sprint Dashboard**
Immagine tratta dal sito http://now.sprint.com/nownetwork/



4. La visualizzazione come strumento per la ricerca e l'analisi


Nuova versione | francesco.roveta@gmail.com | Impostazioni | Account personale | Centro assistenza | Esci
Account personali Analytics: www.blackiwi.com
Segmenti avanzati: Tutte le visite
15/mag/2011 - 14/giu/2011
Grafico per:
20
10
30 mag
6 giu
13 giu
42,51% Frequenza di rimbalzo
00:02:02 Tempo medio sul sito
85,02% % nuove visite
Overlay mappa
Visualizza rapporto
Panoramica dei contenuti
Pagine
Visualizzazio...
% visualizzazioni di pagina
/
299
32,64%
/illuminazione.html
163
17,79%
/contatti.html
112
12,23%



4. La visualizzazione come strumento per la ricerca e l'analisi


Wells Fargo Answer Your Lending
usiness lenders respond to
ers of You're the Read more...
TRAFFIC NOW
Chicago
Kennedy Expressway
GO
STOP
911 CALLS
BEING MADE
0
boingboing
Dalai Lama fails to understand Dalai Lama joke, but is a good sport about it. An Australian newsreader found himself interviewing the Dalai Read more...
PEOPLE FLIRTING ON THE NOW NETWORK™
4112
STICKY NOTES
BEING PRODUCED
55K
46K
37K
28K
18K
9K
0
YOUR SHARE
NEW CASES OF MALARIA
9251
Widget
Screen saver
DAYS UNTIL
CHRISTMAS
193
DAYS UNTIL
FRIDAY
2
CALLS ON SPRINT PHONES
26681794
Video: Larry Mac, Hammond offer congrats to DW. DW's FOX partners reflect on his career, Hall of Fame selection.
Read more...
LISTEN NOW
LET'S STAY FRIENDS
LES SAVY FAV
1. POTS & PANS
2. THE EQUESTRIAN
3. THE YEAR BEFORE...
4. PATTY LEE
5. WHAT WOULD WOLVES DO
6. BRACE YOURSELF
BUY ALBUM @ FRENCHKISS
VIEW BAND @ MYSPACE
NEXTEL DIRECT CONNECT™ CHIRPS NOW
2439884
WORD OF THE MOMENT
dilli
(DIL-ee) noun
Someone or something that is remarkable or unusual.
PLAY NOW
PUSH NOW
uncrate
Workshop Sanction
U.S. MOON
NEXT BUS
Seattle
3:47
Streetcar
Westlake & Denny
PEOPLE STUCK IN ELEVATORS
6
5
4
3
2
1
NES IN THE AIR
ON TIME
DFW TO SJT
TRASH
(million tons)
5650
RECYCLED
(million tons)
182
NASCAR Sprint CUP SERIES
Busch's car fails post-race inspection at Pocono. The No. 18 Toyota driven by Kyle Busch for Joe Gibbs Racing Read more...
HABITABLE PLANETS
1

relazione tra loro e con il contesto in cui si vanno ad inserire. Nel momento in cui il contesto risulta inadeguato si va a compromettere la validità dell'informazione trasmessa dal singolo dato.

Visualizzare un quantità eccessiva di dettagli Una dashboard è una visualizzazione per sua natura orientata alla sintesi. Bisogna sempre essere capaci di mantenere un buon grado di semplificazione per fornire all'utente solo i dati di cui ha realmente bisogno, per non creare un inutile sovraffollamento informativo, spalmando le informazioni in maniera coerente sui diversi livelli di dettaglio.

Scegliere un'unità di misura inappropriata Quando si visualizza un dato, scegliere l'unità di misura con cui questo viene visualizzato è fondamentale per rendere in maniera corretta l'informazione. Anche questo fa parte del processo progettuale, e la scelta va fatta tenendo sempre sott'occhio l'obiettivo che si vuole raggiungere.

Scegliere in maniera inappropriata i supporti di visualizzazione Gli strumenti di visualizzazione a nostra disposizione non sono tutti uguali. Ognuno vanta pregi e difetti in relazione a determinati bisogni di trasmissione dell'informazione. Bisogna sempre saper scegliere quale strumento utilizzare in quale contesto.

Introdurre eccessiva varietà di elementi grafici Una dashboard è per sua definizione un insieme di vari strumenti e visualizzazioni uniti in un'unica schermata per trasmettere una determinata tipologia di informazioni. Si rende necessario quindi saper scegliere quali strumenti visualizzare per evitare da un lato inutili ridondanze di informazione, mentre dall'altro un'eccessiva confusione dal punto di vista percettivo.

Utilizzare visualizzazioni mal progettate Esistono diverse tipologie di visualizzazioni, ognuna delle quali risulta più utile o particolarmente indicata per determinate rappresentazioni. Bisogna sempre saper scegliere quale visualizzazione utilizzare in quale contesto, per sfruttare al massimo i dati in proprio possesso al fine di trasmettere al meglio l'informazione. Inoltre nel momento in cui si progetta una visualizzazione bisogna sempre tener conto dei supporti e dei mezzi tramite i quali questa verrà fruita.

Codificare i dati in maniera imprecisa Bisogna sempre ricordare che chi progetta la visualizzazione funge da mezzo tra il dato di per sé e la trasmissione dell'informazione stessa all'utente. Codificare o trasmettere in maniera errata i dati a quest'ultimo può portare a percezioni errate e distorte degli stessi.

Organizzare male i dati Una dashboard può contenere un gran numero di strumenti e visualizzazioni al suo interno. Saper organizzare al meglio i dati si rende quindi necessario per poter trasmettere al meglio tutte le informazioni necessarie all'utente.

Non evidenziare, o evidenziare in maniera inefficiente dati importanti Come già detto colui che progetta la visualizzazione si pone come tramite tra il dato di per sé e l'informazione data all'utente. Non tutti i dati hanno la stessa importanza, gestire in maniera corretta i dati che necessitano di essere evidenziati fornisce una visualizzazione dove l'informazione è trasmessa in maniera più accurata.

Riempire lo schermo di decorazioni inutili La ricerca estrema dell'estetica fine a se stessa ha come unico risultato quello di distrarre l'attenzione dell'utente dall'informazione vera trasmessa dalla visualizzazione. Inoltre un'eccessiva decorazione porta via spazio utile che potrebbe essere utilizzato per inserire ulteriori dati.

Utilizzare impropriamente il colore Il colore in una visualizzazione è una variabile molto importante. Difatti quest'ultimo trasmette all'utente sensazioni e percezioni che leghiamo intrinsecamente ad alcuni colori. Inoltre vi sono colori più o meno visibili, capaci di attrarre in maniera diversa l'attenzione dell'utente, il colore può essere usato come legante tra due zone diverse di una stessa visualizzazione, e così via.

Progettare una visualizzazione poco attraente Quando una visualizzazione è mal progettata, o semplicemente brutta, l'utente è posto in uno stato mentale non favorevole al suo utilizzo.

Salvando gran parte dei principi espressi fino ad ora, che possono essere riferiti

oltre che alle dashboard anche ad altre tipologie di visualizzazioni, si rende in ogni caso necessario un ragionamento ulteriore sull'attuale utilità di questo tipo di sistemi. Difatti seppur ancora molto utilizzate, e probabilmente in alcuni campi ancora molto utili, le dashboard portano con sé tutta una serie di problematiche legate alla visualizzazione ed alla percezione delle informazioni. Difatti mentre da un lato avere la possibilità di concentrare in un'unica schermata tutte le informazioni utili, il mantenerle divise tra loro, non correlarle, ma posizionarle semplicemente una accanto all'altra, può portare ad una sovrabbondanza disordinata di informazione, capace tanto di di fornire informazioni all'utente quanto di disorientarlo, molto più simile ad un report testuale che ad una visualizzazione progettata. Inoltre in un'organizzazione così frammentata dell'informazione l'utente dispone di tutta una serie di strumenti utili per monitorare singolarmente diverse caratteristiche di un sistema, ma non ha la possibilità di mettere in relazione tra loro informazioni che, anche rifacendosi al modello data-information-knowledge, potrebbero portare alla generazione di nuova conoscenza, e, fattore ancora più importante, uno strumento così strutturato può risultare molto utile come strumento di report, ma essere meno funzionale nei processi di interazione con l'utente e nella scoperta di eventi chiave o casi particolari che possono essere individuati proprio grazie ad una visualizzazione più mirata.

Per tutte queste motivazioni molti sistemi di visualizzazione stanno superando il concetto classico di dashboard, focalizzandosi di volta in volta su di un unica visualizzazione unitaria, contenente anch'essa un grande numero di variabili, ma invece che suddivise tra più sezioni, messe in relazione tra loro su un unico oggetto, al fine di mantenere fissa la concentrazione dell'utente sul singolo, ed al tempo stesso continuando a fornirgli un gran numero di informazioni relative a ciò che sta osservando. Questo processo di concentrazione dell'attenzione dell'utente permette di creare delle visualizzazioni meno frammentarie e confusionarie, dove l'utente mantiene in ogni momento il controllo su ciò che sta osservando, ed, attraverso l'interazione, può spostarsi all'interno della visualizzazione stessa per indagare il fenomeno che sta analizzando, visualizzando di volta in volta soltanto le informazioni che gli sono strettamente necessarie o che l'utente stesso ha richiesto o deciso di osservare.

5 | L'importanza di una collaborazione interdisciplinare

"The in-box culture is dead, so information workers need to learn how to come together on projects spontaneously in real time, instead of handling assignments alone and passing them down the line."

Evan Rosen

Il progetto di tesi qui proposto, descritto in maniera dettagliata nel capitolo sei, pone le sue origini in una proposta di collaborazione tra il DEI (Dipartimento di Elettronica ed Informazione del Politecnico di Milano) ed il Density Design (laboratorio di ricerca del dipartimento INDACO del Politecnico di Milano). Il punto di incontro tra queste due realtà è stata la necessità da parte del DEI di sviluppare un'applicazione che utilizzasse la visualizzazione come mezzo per analizzare, ricercare e comunicare dati relativi al campo della sicurezza informatica. Rientrando nei programmi del progetto WOMBAT (meglio descritto nei prossimi capitoli), il DEI ha iniziato a lavorare sui dati forniti dal database FIRE nel tentativo di sviluppare da questi ultimi nuovi algoritmi di detection delle attività malevole, al fine di ottenere nuove informazioni e quindi nuova conoscenza su questo fenomeno. Questo lavoro ha trovato una risposta nel progetto di tesi di Luca Di Mario [1], tesi di laurea in ingegneria informatica, che ha iniziato a sviluppare alcuni algoritmi per trattare i dati grezzi provenienti dal database. Mentre Luca si occupava della parte di analisi sui dati, io ho iniziato una ricerca dal punto di vista della visualizzazione, con l'obiettivo comune di creare un'applicazione online, che si autoaggiornasse giornalmente coi nuovi dati provenienti dalla rete ed accessibile a tutti gratuitamente, che attraverso la visualizzazione

[1] "BURN: Baring Unknown Rogue Networks - Studio e sviluppo di un'applicazione per l'analisi del comportamento delle reti malevole". Luca Di Mario, Tesi di laurea magistrale in ingegneria informatica.

non solo comunicasse all'utente finale dati e informazioni, aumentandone la consapevolezza relativamente all'argomento trattato, ma permettesse anche di fornire i mezzi per un'analisi interattiva, fungendo quindi sia come mezzo di conoscenza, diffusione e comunicazione, sia come strumento a servizio di esperti e ricercatori del settore. Da questa collaborazione è nato BURN, un sistema per la visualizzazione e l'analisi del comportamento dei network malevoli.

5.1 Sicurezza informatica e rappresentazione visiva

Questo progetto di tesi nasce dall'unione e dalla collaborazione di due realtà, appartenenti a diverse facoltà, che proprio grazie alle loro diverse competenze hanno permesso lo sviluppo di un lavoro completo sia dal punto di vista ingegneristico dello sviluppo e della ricerca nel campo informatico, sia dal punto di vista del design della comunicazione.

Il VPlab (laboratorio di Valutazione delle Prestazioni e Affidabilità) [2] è una realtà interna al DEI (il Dipartimento di Elettronica e Informazione del Politecnico di Milano) [3] e si occupa di valutazione delle prestazioni, affidabilità e sicurezza dei sistemi informatici come server, reti e networks, macchine virtuali, e così via.

Il DensityDesign [4] è un laboratorio di ricerca interno al dipartimento INDACO del Politecnico di Milano. La sua attività di ricerca si focalizza sulla rappresentazione visiva di fenomeni e dati complessi. Obiettivo di questo tipo di ricerca è quello di esplorare le potenzialità dell'information visualization e dell'information design e produrre artefatti visivi innovativi e affascinanti, che permettano di scoprire, descrivere e narrare le connessioni nascoste in qualsiasi sistema o fenomeno complesso.

[2] http://www.vplab.elet.polimi.it/
[3] http://www.dei.polimi.it/
[4] http://www.densitydesign.org/

5.2 Inserirsi in un progetto europeo: WOMBAT e FIRE

Come accennato nell'introduzione al capitolo sei, il progetto presentato in questa tesi si inserisce all'interno di un contesto europeo, ed in particolar modo all'interno del progetto WOMBAT. Come partecipante attivo al progetto, il DEI aveva come ultimo obiettivo quello di migliorare la resa pubblica dei dati grezzi raccolti dal database di FIRE, cosa che fino a questo momento era fatta solamente attraverso il sito maliciousnetworks.org. Proprio per migliorare tale strumento, il nostro progetto di tesi si è inserito all'interno di questo programma di ricerca, con l'obiettivo finale di produrre uno strumento fruibile online che affiancasse il sito già esistente e fornisse un maggiore contributo sia dal punto di vista della comunicazione sia da quello dell'analisi e della ricerca.

5.2.1 WOMBAT, un progetto lungo tre anni

Il progetto WOMBAT, Wombat Worldwide Observators of Malicious Behaviours and Attack Threats [5], nato nel 2008 e conclusosi nei primi mesi del 2011, è un progetto di ricerca finanziato dalla Commissione Europea ed istituito da un consorzio di organizzazioni di ricerca tecnologica, come il Politecnico di Milano, l'Università di Amsterdam Vrije, l'Università Tecnica di Vienna e a cui partecipano società internazionali del settore sicurezza Internet come Symantec.

Questo progetto nasce in risposta alla sempre maggiore necessità di combattere tutto quel mondo di attacchi di tipo criminale che giornalmente si verificano sulla rete internet, e la necessità di una collaborazione di tale portata è dettata da due considerazioni principali: per prima cosa oggigiorno i malware vengono sempre più progettati su misura delle più moderne tecniche di difesa, sia tecnologiche, che economiche, che sociali, e per venire a capo di questa situazione si rende sempre più utile un approccio mirato a comprendere il fenomeno nella sua interezza piuttosto che sviluppare linee di difesa per i singoli attacchi; in secondo luogo negli ultimi anni si è creata una vera e propria rete di organizzazioni criminali che agiscono secondo logiche economiche e di profitto ormai ben avviate. Per queste ragioni si è reso necessario uscire dall'isolamento e dare

[5] http://wombatproject.eu

una svolta al campo della sicurezza informatica tentando un approccio di tipo condiviso che permettesse un'analisi più generalizzata del fenomeno nel suo insieme, nel tentativo di trovare dei metodi per comprenderne funzionamento ed andamenti e prendere le giuste contromisure.

Il progetto è stato organizzato in tre fasi:

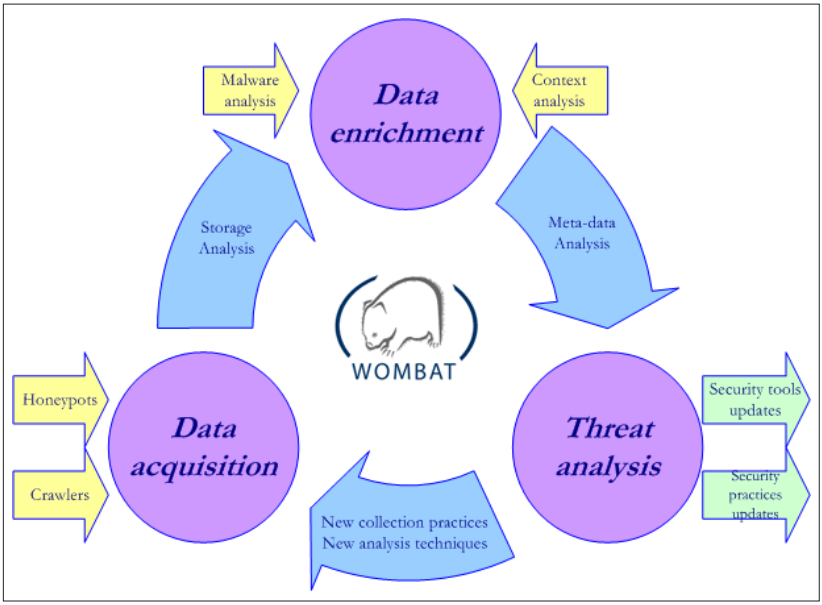
Fase 1 Raccogliere in tempo reale svariati set di dati grezzi relativi al campo della sicurezza informatica. In questo modo si potranno sfruttare gli sforzi condivisi di raccolta dati tra tutti i partner e le organizzazioni che fanno parte del progetto. In questa prima fase ci si concentrerà sullo sfruttamento di strumenti già esistenti e sull'esplorazione di uno sviluppo di strumenti dedicati al wireless (wifi, RFID, bluetooth) ed alle reti.

Fase 2 Arricchire i dati grezzi raccolti nella prima fase per mezzo di varie tecniche di analisi, con l'obiettivo di formalizzare le informazioni sulle minacce.

Fase 3 Analisi delle minacce. WOMBAT si baserà sulle informazioni così raccolte per fornire una più accurata *root cause analysis*. Questo fornirà una maggiore identificazione e comprensione dei fenomeni presi in esame, che potrebbe condurre alla creazione di sistemi semiautomatici in grado di risalire agli autori e distributori dei programmi malware, nonché identificare trend futuri per prevenire la loro diffusione sistematica.

Al termine di queste tre fasi le conoscenze acquisite grazie a questo sforzo collettivo verranno condivise con tutti gli attori ed esperti di sicurezza informatica, permettendo loro, oltre che di condividere nuove conoscenze, di poter prendere le giuste decisioni in ambito di sicurezza e di concentrare la loro attenzione sulle attività più pericolose. Inoltre il progetto si pone anche come obiettivo quello di aumentare la consapevolezza del cittadino europeo nei confronti di questa tipologia di minacce, al fine di sensibilizzarlo nei confronti

▶ **fig 28 | Ciclo delle fasi del progetto WOMBAT**
Immagine tratta dal sito http://wombat-project.eu/wombat-project-description.html



dell'argomento, per aumentare la sicurezza collettiva.

"Dal luglio 2008 al giugno 2009 abbiamo individuato i nuovi trend della cybercriminalità: tra questi, oltre alla ricerca di dati su carte di credito, furto d'identità ed altro, abbiamo visto in crescita esponenziale il fenomeno del sequestro informatico dei computer di ignari utenti: in pratica attraverso finti programmi che annunciano la presenza di virus negli elaboratori, sfruttando la paura e l'ignoranza del titolare, si ottiene il pagamento di una cifra variabile per ripulire il Pc, istallando un software che in realtà, poi, spia tutte le attività dell'utente stesso. Abbiamo individuato ben 43 milioni di casi in tutto il mondo e all'utente questi casi sono costati da 50/60 centesimi a computer a un dollaro. E' facile calcolare quali guadagni ci possano essere dietro questo fenomeno che fa leva su tecniche di social engineering, sfruttando la paura o l'ansia dell'utente [...] Sono dati, questi, raccolti proprio dal sistema WOMBAT. Abbiamo costituito un vero e proprio database globale che registra gli attacchi informatici e i trend di sviluppo futuro. Questi dati grezzi sono poi stati elaborati per individuare gli algoritmi necessari per avviare procedure pseudoautomatiche di individuazione delle

direttrici di attacco e per fornire a tutti gli operatori della sicurezza on line gli strumenti d'analisi per aggiornare i propri sistemi. In questo contesto, poi, ci siamo resi conto che le direttrici d'attacco si andavano spostando dai sistemi "zombie" (Pc governati in remoto all'insaputa degli ignari utenti) ai sistemi di social network, con lo stesso principio dello sfruttare l'ansia e la paura degli utenti o la loro curiosità. Ma il vero obiettivo finale del sistema WOMBAT è quello di identificare i modi più idonei per sottrarre alla cybercriminalità il suo obiettivo principale: fare soldi. In questo senso, individuando con anticipo i trend futuri e le direttrici di attacco, è possibile impedire che i flussi di programmi pericolosi arrivino all'utente finale. Ci vorrà ancora del tempo per questo, ma contiamo di avere risultati concreti entro il 2011." [6]

5.2.2 FIRE e malicious networks

Come già spiegato nel capitolo tre, FIRE è un sistema che mira ad identificare ed esporre tutte quelle organizzazioni ed ISPs (Internet Service Providers) che dimostrano un persistente atteggiamento malevolo sulla rete. L'obiettivo è quello di identificare tutti quei network che sono costantemente implicati in attività illecite. Sviluppato all'interno del progetto europeo WOMBAT, FIRE serve a monitorare giornalmente l'attività malevola presente sulla rete e ne raccoglie i dati, inserendoli in un database. Inoltre, attraverso il sito maliciousnetworks.org, alcuni di questi dati vengono resi pubblici, a disposizione di chiunque voglia prenderne visione.

Oltre a questo, anche altri database sono stati sviluppati sempre all'interno del progetto WOMBAT (Harmur, Forth, Anubis, Wepawet, SGNet), ognuno dei quali con dati e caratteristiche differenti. In tutti questi casi però l'attenzione verso una chiara esposizione dei dati nei confronti di un pubblico di non soli esperti del settore non è sempre stata sufficiente. Lo sforzo più grande da questo punto di vista è stato fatto proprio da FIRE attraverso il sito maliciousnetworks.org, ma anche in questo caso le informazioni proposte sul web sono piuttosto scarne e non sempre di facile accesso, soprattutto per un'utenza senza specifiche competenze nell'argomento. Inoltre i dati presenti

[6] Antonio Forzieri e Stefano Zanero. "Wombat, l'Europa contro il cybercrime. Università e aziende studiano i pericoli futuri" articolo di Claudio Gerino, su "la Repubblica.it". 23 novembre 2009

risultano a volte difficilmente accessibili ed un loro rapido confronto per l'analisi di alcuni specifici fenomeni non sempre è di facile attuazione.

Col nostro progetto di tesi, io e Luca Di Mario, ci siamo inseriti proprio in questo contesto, con l'obiettivo sia di analizzare e studiare i dati grezzi presenti nel database di FIRE per trovare nuove relazioni e connessioni che potessero generare nuova informazione, sia di costruire un modello di visualizzazione di questi stessi dati che avesse la duplice funzione di aiutare la resa e la comunicazione dei risultati ed al tempo stesso funzionare come strumento di analisi per ricercatori ed esperti del settore.

5.3 I dati, un punto di partenza per creare nuova informazione

Come spiegato il nostro lavoro si è basato sui dati raccolti dal progetto FIRE e inseriti nel suo database. Prima fase del progetto è stata quindi quella di andare ad analizzare quali dati grezzi erano in nostro possesso e a cosa ognuno di questi dati si riferiva, con eventuali connessioni già esistenti tra loro. Successivamente abbiamo analizzato come FIRE già utilizzasse questi dati, in che modo li raggruppasse e come alcuni interagissero tra loro, sia a livello di database sia a livello di sito online (maliciousnetworks.org). Inoltre tramite il sito abbiamo potuto osservare come alcuni di questi dati venissero già tradotti in grafici e visualizzazioni, comparando queste ultime con le visualizzazioni utilizzate anche da altri sistemi, per avere un confronto con lo stato attuale dell'arte nel settore. Solo in seguito a questa fase di studio e analisi abbiamo iniziato a costruire le prime idee ed ipotesi sul come fosse utile poter rimaneggiare i dati grezzi per generare informazioni non ancora presenti in questo sistema, o per migliorare quelle già presenti.

5.3.1 I dati grezzi: come FIRE li raccoglie e li organizza

Il database di FIRE si preoccupa di raccogliere al suo interno informazioni riguardanti quattro tipologie di attacchi informatici (c&c, malware, phishing, spam), prendendo nota della data in cui ogni attacco viene individuato ed in quale AS, autonomous system (gruppo di router e reti, all'interno della rete internet, sotto il controllo di una singola e ben definita autorità amministrativa) questo attacco

viene registrato. Inoltre i dati raccolti da FIRE vengono già processati in maniera tale da ottenere alcune informazioni aggiuntive, pesate in base a parametri che servono per normalizzare le informazioni ottenute (vedi capitolo 3.1.2 EMBER).

La struttura di base del database è composta da sette tabelle all'interno delle quali le informazioni vengono registrate ed organizzate:

'asn_cache' si tratta di una raccolta di tutti gli AS registrati dal sistema;

'bots_incidents' contiene l'elenco di tutti gli attacchi C&C registrati;

'download_incidents' come la precedente ma per gli attacchi ti tipo malware;

'phish_incidents' attacchi di phishing;

'spam_incidents' attacchi di spam;

'flagged_asns_incidents' in questa tabella sono presenti tutti gli attacchi, con anche alcune informazioni non grezze, come ad esempio un valore di score basato sul livello di malevolenza delle reti, ed un rank che serve a classificare le reti sempre in base alla loro malevolenza;

'flagged_asns_incidents_nospam' questa tabella è in tutto simile alla precedente solo che nel calcolo del valore di score non tiene conto delle attività di spam (questo poiché a volte le attività di spam risultano di molto maggiori rispetto a tutte le altre).

`asn_cache`
`id`
`ip`
`asn`
`asn_desc`
`route`
`ts`

`bots_incidents`
`id`
`ip`
`asn`
`asn_desc`
`bot_type`
`route`
`country`
`latitude`
`longitude`
`date_added`

`download_incidents`
`id`
`ip`
`asn`
`asn_desc`
`route`
`country`
`latitude`
`longitude`
`date_added`

`phish_incidents`
`id`
`ip`
`asn`
`asn_desc`
`route`
`country`
`latitude`
`longitude`
`date_added`

`spam_incidents`
`id`
`ip`
`asn`
`asn_desc`
`route`
`country`
`latitude`
`longitude`
`date_added`

`id` è il numero che identifica la riga nel database, non rappresenta un dato, è una riga di servizio; `ip` numero che identifica univocamente i dispositivi collegati con una rete informatica che utilizza lo standard Internet Protocol, tra le altre cose dall'ip è possibile ricavare la dimensione della rete a cui appartiene; `asn` e `asn_desc` sono il numero identificativo ed il nome di ogni Autonomous System presente nella rete; `route` è il percorso informatico che i pacchetti devono fare per raggiungere la rete; `ts` è il "time stamp", registra la data in cui una riga viene aggiunta nel database; `bot_type` contiene le informazioni circa la tipologia di attacco c&c portato; `country` è lo stato a cui appartiene l'AS in

`flagged_asns_incidents`	`flagged_asns_incidents_nospam`
`id`	`id`
`asn`	`asn`
`asn_desc`	`asn_desc`
`score`	`score`
`base_score`	`base_score`
`rank`	`rank`
`route`	`route`
`country`	`country`
`latitude`	`latitude`
`longitude`	`longitude`
`bots`	`bots`
`phishs`	`phishs`
`spams`	`spams`
`malservers`	`malservers`
`date_added`	`date_added`
`asn_size`	`asn_size`

cui viene registrata attività malevola; `latitude` e `longitude` sono le coordinate di registrazione dell'AS; `date_added` svolge la stessa funzione del "time stamp";

In aggiunta a questi dati FIRE nelle ultime due tabelle presentate inserisce anche ulteriori righe che tengono conto del numero di attacchi registrati suddivisi per tipologie `bots` (C&C), `phishs` (phishing), `spams` (spam), `malservers` (malware), più una riga dedicata alla dimensione della rete `asn_size`. Inoltre sono presenti anche tre contatori, che rappresentano informazioni create da FIRE utilizzando i dati precedentemente elencati: `score`e `base_score` sono un punteggio di malevolenza dato ad ogni singolo AS e calcolato normalizzando il numero di attacchi registrati all'interno di ogni singola rete con la sua dimensione; `rank` è la "posizione in classifica" di ogni AS giorno per giorno, posizione che si basa sul valore di score associato alla rete.

Analizzando queste tabelle si può quindi fare un primo ragionamento sul come FIRE organizza i suoi dati. Esistono due principali modalità di indicizzazione dei dati: un'indicizzazione basata sulla localizzazione spazio/temporale degli Autonomous System, ed un'indicizzazione basata su alcuni contatori che mirano a creare una classificazione, sempre degli Autonomous System. Per ogni AS conosciamo il numero di attività malevole registrate al suo interno in un determinato periodo, con una loro suddivisione nelle quattro tipologie di attività precedentemente elencate. Inoltre abbiamo alcuni dati aggiuntivi, come ad esempio l'ip o l'asn_size, che possono essere utili per estrapolare ulteriori informazioni.

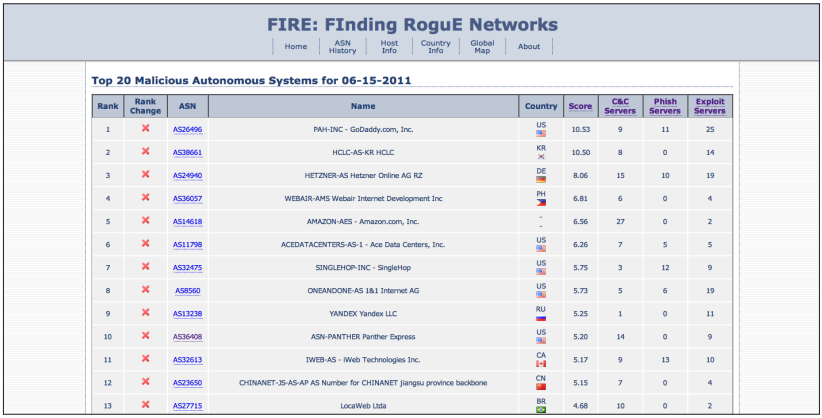
Fatta questa prima analisi sui dati siamo passati ad un'osservazione su come questi venissero trasmessi al di fuori del database.

5.3.2 L'attuale resa grafica: maliciousnetworks.org

L'unico modo per accedere ai dati raccolti da FIRE per tutti coloro che non sono in possesso dei permessi per utilizzare direttamente il database, è interfacciarsi ad essi attraverso il sito maliciousnetworks.org. Questo portale si struttura in sei sezioni, e riporta su una piattaforma gratuitamente accessibile da chiunque una buona parte dei dati raccolti (non tutti).

Il sito si compone di sei sezioni, riportate qui di seguito:

Home Prima pagina in cui si atterra accedendo al sito, contiene una tabella dove vengono elencati i 20 Autonomous System considerati più malevoli. I dati si riferiscono sempre al giorno precedente la visualizzazione, in quanto il database viene aggiornato ogni mattina coi dati del giorno precedente. Una prima limitazione di questo sistema risiede proprio in questa tabella, dove è possibile consultare i dati solamente dei primi 20 AS classificati, e solamente riferiti al giorno precedente. Inoltre in questa tabella mancano i dati relativi ad attività di spam, e se la si riordina in base ad una delle tre tipologie di attività presenti invece che in base allo score (ordine di default), la visualizzazione dei soli primi 20 risultati risulta ancor più limitante. Mancano anche dei riferimenti che aiutino nella comprensione un utente non esperto nell'argomento.

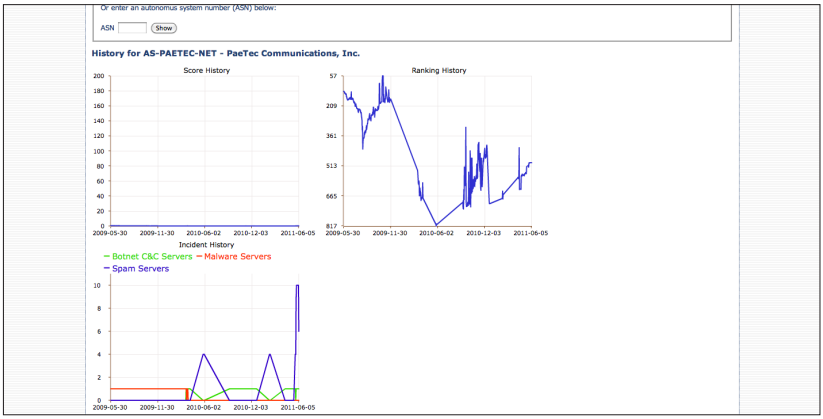


FIRE: FInding RoguE Networks

Home | ASN History | Host Info | Country Info | Global Map | About

Top 20 Malicious Autonomous Systems for 06-15-2011

Rank	Rank Change	ASN	Name	Country	Score	C&C Servers	Phish Servers	Exploit Servers
1	✖	AS26496	PAH-INC - GoDaddy.com, Inc.	US	10.53	9	11	25
2	✖	AS38661	HCLC-AS-KR HCLC	KR	10.50	8	0	14
3	✖	AS24940	HETZNER-AS Hetzner Online AG RZ	DE	8.06	15	10	19
4	✖	AS36057	WEBAIR-AMS Webair Internet Development Inc	PH	6.81	6	0	4
5	✖	AS14618	AMAZON-AES - Amazon.com, Inc.	-	6.56	27	0	2
6	✖	AS11798	ACEDATACENTERS-AS-1 - Ace Data Centers, Inc.	US	6.26	7	5	5
7	✖	AS32475	SINGLEHOP-INC - SingleHop	US	5.75	3	12	9
8	✖	AS8560	ONEANDONE-AS 1&1 Internet AG	US	5.73	5	6	19
9	✖	AS13238	YANDEX Yandex LLC	RU	5.25	1	0	11
10	✖	AS36408	ASN-PANTHER Panther Express	US	5.20	14	0	9
11	✖	AS32613	IWEB-AS - iWeb Technologies Inc.	CA	5.17	9	13	10
12	✖	AS23650	CHINANET-JS-AS-AP AS Number for CHINANET Jiangsu province backbone	CN	5.15	7	0	4
13	✖	AS27715	LocaWeb Ltda	BR	4.68	10	0	2



Or enter an autonomus system number (ASN) below:

ASN Show

History for AS-PAETEC-NET - PaeTec Communications, Inc.

Score History

Ranking History

Incident History

Botnet C&C Servers Malware Servers Spam Servers

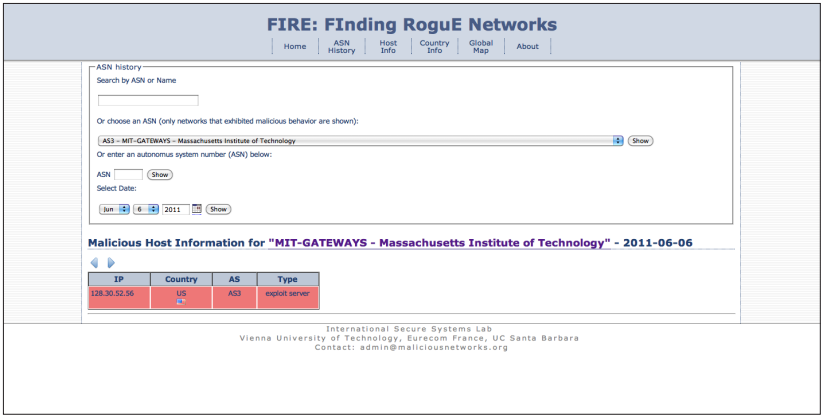
▲ **fig 29a-b | maliciousnetworks.org, Home e ASN History**
Immagini tratte dal sito http://maliciousnetworks.org/

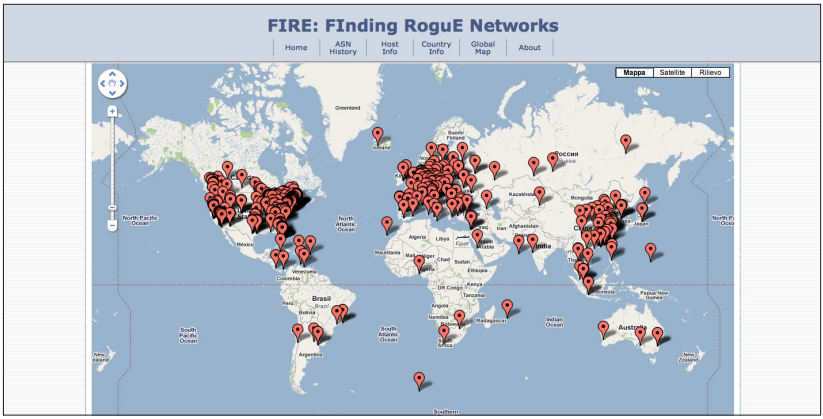
ASN History In questa sezione è possibile visionare, attraverso l'ausilio di alcuni grafici, la storia relativa ai singoli AS. Scegliendo di visualizzare la storia di un AS appaiono tre grafici: uno che mostra il suo punteggio di score nel corso del tempo, uno che mostra il suo rank nel corso del tempo, ed uno che mostra l'andamento su questa rete delle diverse tipologie di attività malevole. Anche in questo caso ci si scontra con alcune problematiche legate alla comprensione dei dati visualizzati. Innanzi tutto la dimensione e la scala dei grafici rendono difficile una chiara visione degli andamenti, visualizzati con una grana troppo fine in una dimensione molto limitata. Si hanno in questo modo dei grafici molto densi, dove, nonostante la presenza di un tooltip interattivo al passaggio del mouse, è difficile risalire ai singoli dati visualizzati. Inoltre nel grafico dove vengono visualizzate le diverse tipologie di attività compare in questo caso l'attività di spam, e ciò che nella home era indicato come Exploit Server diviene qui Malware Server, creando problemi di coerenza di denominazione, che come prima, non permettono una chiara comprensione ad un utente non esperto.

Host Info In questa sezione viene fornito un dettaglio sui singoli AS. Giorno per giorno, attraverso una tabella, si può andare a visionare quali server sono stati individuati come malevoli all'interno dell'AS, e a quale tipologia di attività malevola sono stati abbinati.

Country Info Similmente alla Home, questa sezione contiene una tabella dove vengono elencati i 20 stati che hanno registrato nella giornata precedente il maggior numero di server malevoli. Vi è anche uno score abbinato allo stato ed una suddivisione tra il numero di server appartenenti alle quattro tipologie di attività registrate.

Global Map Questa sezione presenta una google map dove sono posizionati gli AS registrati come malevoli. Anche questa visualizzazione presenta diverse problematiche. Prima di tutto la visualizzazione a livello mondiale risulta, a causa della grande quantità di puntatori presenti anche a questo livello di zoom, un insieme confusionario ed indistinto di AS, dove è difficile da parte dell'utente riuscire anche solo a capire a che stato appartengono od a selezionarne uno. Zoommando sulla mappa questo disagio viene mitigato, ma





▲ **fig 30a-b | maliciousnetworks.org, Host Info e Global Map**
Immagini tratte dal sito http://maliciousnetworks.org/

continuando nello zoom si arriva ad un problema di tipo opposto. Difatti è possibile zoommare fino al massimo livello consentito dalla google map. In questo modo gli AS vengono posizionati in maniera fin troppo precisa sulla mappa, portando a tutte le problematiche e conseguenze di cui parleremo nel capitolo 6.2.2 (Geographical Map).

About E' l'about del sito. Contiene una breve descrizione dell'argomento trattato e presenta il database di FIRE, con tutti i riferimenti utili ad un suo approfondimento.

Osservando nell'insieme questo strumento si può quindi notare come, nonostante sia una buona fonte di dati per tutti gli operatori del settore, soffra di non poche problematiche, sia dal punto di vista della completezza e coerenza dei dati forniti, sia per una mancanza di possibilità di interazione tra le diverse modalità di report presenti, sia per alcune problematiche tecniche strettamente legate alle modalità di visualizzazione.

5.4 La nostra proposta

Combinando i dati grezzi presi dal database di FIRE con la loro comunicazione al pubblico, riportata sul sito maliciousnetworks.org, abbiamo iniziato a ragionare sul come poter migliorare questo servizio e su quali nuovi dati potevano essere studiati, combinati ed infine visualizzati. Abbiamo quindi lavorato in due direzioni: migliorare la resa dei dati già esistenti e creare nuovi dati, e soprattutto nuova informazione, attraverso i dati grezzi in nostro possesso.

5.4.1 Migliorare la resa dei dati già esistenti

Per migliorare la resa dei dati già esistenti il punto di partenza è stato il sito maliciousnetworks.org, con un'analisi della sua struttura e soprattutto delle sue limitazioni dal punto di vista della comunicazione al pubblico.
L'obiettivo del nostro progetto è di fatti quello di creare uno strumento sì utile agli esperti del settore nei processi di analisi, report e ricerca delle attività malevole, ma anche utile all'avvicinamento al problema di un ipotetico utente senza conoscenze nel settore, questo poiché come abbiamo visto nei capitoli introduttivi

di questa tesi comunicare ad un più ampio bacino di persone significa esporre maggiormente al pubblico queste attività criminali, il che spesso porta ad una diminuzione delle stesse, ed aumentare la consapevolezza pubblica nei confronti di questa tipologia di minacce. Per la combinazione di queste due motivazioni si è reso necessario lo studio di uno strumento che fosse polivalente, che fornisse diversi livelli di complessità ed utilizzo, e che fosse accessibile a tutti tramite la rete. BURN è stato quindi progettato per essere un sistema online, liberamente fruibile sul web, che affianchi l'attuale sito maliciousnetworks.org.

Osservando i dati presenti nel database e le informazioni visualizzate su maliciousnetworks.org si può notare che non tutti i dati presenti vengono sfruttati a pieno, alcuni non vengono riportati del tutto, e le visualizzazioni spesso si riducono a tabelle e grafici, che possono essere sì utili come report, ma non come strumento attivo di analisi. Inoltre analizzando il fenomeno ci siamo resi conto che uno degli aspetti forse più interessanti è osservare come questo vari nel tempo, per analizzare trend, comportamenti ed abitudini di chi pratica questa tipologia di attività illegali. Tramite il sito attuale è possibile visualizzare i dati relativi solamente al giorno precedente, mentre sarebbe molto più interessante avere la possibilità di scegliere quale giorno andare ad analizzare, o ancora meglio quale periodo di tempo prendere in considerazione. Alcuni fenomeni difatti, come ad esempio l'attivazione o disattivazione di numerosi server all'interno di uno stesso AS, possono richiedere anche diversi giorni.

Tenendo in considerazione quindi le problematiche riscontrate sull'attuale sito, caratteristiche e funzionalità di altri strumenti di visualizzazione che si occupano dello stesso fenomeno, e i principi di progettazione legati ad una buona comunicazione e visualizzazione, alcuni dei quali sono stati introdotti nel capitolo 4 "La visualizzazione come strumento per la ricerca e l'analisi", nelle fasi di progettazione di BURN abbiamo tenuto in considerazione quanto segue:
- l'utilità di una timeline che permettesse all'utente di selezionare a piacimento il periodo di tempo da andare a visualizzare ed analizzare
- la presenza di più visualizzazioni, una per ogni tipologia di informazione visualizzata, per concentrare al meglio l'attenzione dell'utente sull'informazione richiesta
- la necessità di un sistema di interazione che permettesse un rapido passaggio

da una visualizzazione all'altra, senza perdita di informazioni nei vari passaggi, a supporto di una navigazione delle informazioni il più possibile completa
- la costruzione di un primo livello di interazione contenente le informazioni generali adatte ad un pubblico più ampio, e di un secondo livello di dettaglio più utile ad un pubblico di esperti del settore
- l'importanza di una corretta comunicazione delle caratteristiche fisiche dei singoli AS (dimensione, tipologia di server al loro interno, etc etc), e di una corretta comunicazione geografica del fenomeno
- la possibilità di tenere traccia delle proprie ricerche
In aggiunta a tutto questo abbiamo lavorato alla creazione di nuova informazione.

5.4.2 Generare nuova informazione

Secondo punto del nostro progetto di tesi è stato quello di creare, attraverso lo studio dei dati grezzi presi dal database di FIRE, nuovi dati e quindi nuova informazione. Per fare questo abbiamo iniziato a ragionare su quali nuove informazioni potessero essere estratte dai dati presenti, e dopo aver analizzato svariate ipotesi abbiamo concentrato la nostra attenzione su due punti:
- osservare gli incrementi e i decrementi importanti di attività all'interno di un AS, fino a stabilire quando si può parlare di shutdown (evento che si verifica quando il decremento di attività all'interno dell'AS è molto importante e si verifica in un lasso di tempo breve, si ha una vera e propria caduta della curva di attività, che può anche corrispondere ad uno spegnimento dei server malevoli e quindi ad una cessazione completa dell'attività stessa), o di turn on (quando al contrario si registra una crescita improvvisa e molto marcata dell'attività all'interno di un AS);
- osservare la permanenza di un server registrato come malevolo all'interno di uno stesso AS, il che può essere indice di un Autonomous System che non si preoccupa di prendere provvedimenti a riguardo, o non li prende in maniera adeguata.

Sulla base di queste riflessioni si è iniziato a lavorare in maniera euristica sui dai grezzi, attraverso la creazione di nuovi algoritmi che permettessero:
- un calcolo mirato a individuare gli spostamenti di attività malevola da una rete all'altra (attraverso un'analisi degli incrementi e decrementi di attività);
- un calcolo del livello di tolleranza delle reti che permettono ai server malevoli di perpetrare indisturbati le loro attività per lunghi periodi di tempo.

Analisi delle migrazioni di attività

Abbiamo sviluppato una semplice euristica per riconoscere segni di possibili migrazioni di attività tra diversi AS, tra quelli che abbiamo chiamato AS sorgente e AS destinatario. L'assunto fondamentale dell'euristica è che queste migrazioni sono composte da una fase di shutdown (osservabile nell'attività dell'AS sorgente) seguita da una fase di attivazione (osservabile nell'attività dell'AS destinatario). La fase di shutdown si può avere per svariate motivazioni: coloro che perpetravano l'attività malevola hanno disattivato i loro server presenti nell'AS sorgente prima di trasferirsi nell'AS destinatario, oppure l'AS sorgente ha tagliato la connessione obbligando i criminali a spostare le loro attività su di un altro AS, e così via. Quando il numero dei server registrati come malevoli all'interno di un AS cala improvvisamente, e in un altro AS viene registrato un improvviso incremento di attività malevola corrispondente, questo viene individuato dal sistema come indice di una possibile migrazione.
Poiché le migrazioni possono essere di svariate tipologie e c'è anche la possibilità di migrazioni verso destinazioni multiple il nostro sistema contiene anche uno score di compatibilità che serve a individuare le migrazioni che hanno un maggiore corrispondenza tra l'attività di partenza e quella di destinazione, e quindi un più alto tasso di probabilità di essere una reale migrazione della stessa attività o gruppo di persone.

Analisi della tolleranza di un AS

Alcuni Autonomous System ospitano al loro interno server riconosciuti come malevoli, ai quali viene permesso di continuare indisturbati nelle loro attività anche per periodi di tempo molto lunghi. Poiché la responsabilità delle attività che vengono svolte all'interno di una rete è anche di colui dal quale la rete viene amministrata, e quindi colui che gestisce l'AS, quegli Autonomous System che non prendono le adeguate contromisure per limitare la crescita di attività malevola al loro interno vengono indicati come tolleranti. Per creare un indice di tolleranza il sistema monitora giornalmente l'attività dei diversi server all'interno di un AS, e quando questa attività supera una soglia stabilita, l'AS viene marcato.

Entrambe queste informazioni, per la loro importanza dal punto di vista dell'analisi e della ricerca, godono di una visualizzazione dedicata all'interno di

BURN, visualizzazioni descritte all'interno del capitolo 6.3 "Secondo livello di analisi: l'autonomous system view". Per un maggiore approfondimento sulla costruzione di calcoli ed algoritmi, sopra descritti in maniera molto semplificata, fare riferimento alla tesi di Luca Di Mario [1].

[1] "BURN: Baring Unknown Rogue Networks - Studio e sviluppo di un'applicazione per l'analisi del comportamento delle reti malevole". Luca Di Mario, Tesi di laurea magistrale in ingegneria informatica.

6 | BURN, Baring Unknown Rogue Networks

"Complicare è facile, semplificare è difficile. Per complicare basta aggiungere, tutto quello che si vuole: colori, forme, azioni, decorazioni, personaggi, ambienti pieni di cose. Tutti sono capaci di complicare. Pochi sono capaci di semplificare."

Bruno Munari

Nato dalla collaborazione descritta nel capitolo precedente, il nostro progetto di tesi, BURN, si propone come un'applicazione online, liberamente fruibile, che contiene al suo interno tutti i dati provenienti dal database di FIRE più altri calcolati appositamente per sviluppare alcune sezioni aggiuntive. Sfruttando i principi dell'information e data visualization questo progetto di tesi propone un sistema che permetta di visualizzare giorno per giorno la situazione delle attività malevole sulla rete, visualizzando sia dati generici inerenti i vari Autonomous System rilevati come portatori di attività malevola, come ad esempio loro posizione, numero e tipologia di attacchi e responsabilità, sia dati più specifici, come gli IP appartenenti ad ogni singolo AS. Inoltre, a tutti i dati presi da FIRE, BURN aggiunge alcuni calcoli algoritmici utili per osservare il livello di tolleranza che i vari AS hanno nei confronti di queste attività malevole, e per monitorare gli spostamenti di attività da una rete all'altra. Il nostro progetto si inserisce all'interno di un vasto panorama di strumenti e programmi utili a questo scopo, dei quali però solo una minima parte si preoccupa di sfruttare i mezzi forniti dalla visualizzazione per rendere al meglio i dati in proprio possesso. Con BURN abbiamo contribuito a spingere verso quella collaborazione interdisciplinare indispensabile per creare progetti completi e funzionali sotto tutti i loro punti di vista, che in futuro potrebbe portare a modificare il panorama degli strumenti ora presenti sul mercato.

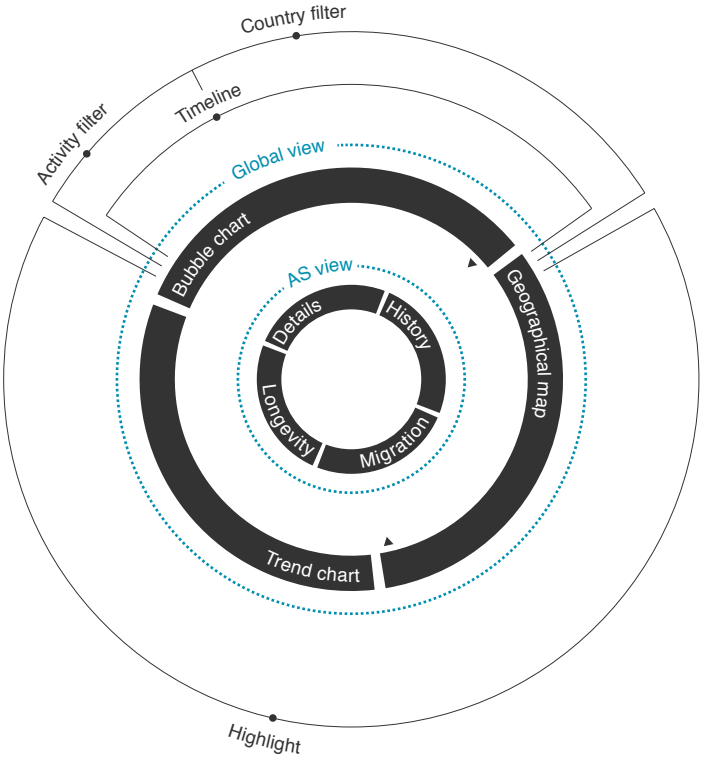
6.1 Macrostruttura del sistema

BURN si struttura come un sistema basato su due livelli di analisi: un primo livello, chiamato *global view*, che visualizza e serve ad analizzare la situazione a livello globale, di tutti gli Autonomous System presenti sul pianeta; ed un secondo livello, chiamato *autonomous system view*, che va ad indagare il singolo Autonomous System nel dettaglio, permettendone un analisi più precisa, tecnica e dettagliata. Il primo livello di analisi, progettato tenendo conto dei bisogni di un utente non esperto, mira a fornire in maniera chiara ed intuitiva i mezzi per un'analisi generale del fenomeno. Il secondo livello di analisi invece è stato progettato sui bisogni di quello che può essere un utente esperto del settore, od un ricercatore, che mira ad analizzare il problema più nello specifico. La comunicazione tra questi due macro-livelli è sempre garantita da alcuni strumenti e filtri, che permettono di tenere traccia delle ricerche effettuate, modificare alcuni parametri di ricerca, e così via.

Altro elemento di fondamentale importanza presente in BURN è la timeline. Posizionata nella parte inferiore della schermata, permette di selezionare il periodo che si desidera analizzare, da un minimo di un giorno ad un massimo di tutti i giorni presenti nei dati registrati dal database. Questa funzionalità è molto importante in quanto raramente o quasi mai nei sistemi esistenti attualmente sul mercato vi è la possibilità di scegliere il periodo di tempo da andare ad analizzare. Spesso i sistemi in commercio si limitano a presentare la situazione degli ultimi giorni, o un'analisi suddivisa per anni. Con BURN abbiamo voluto dare invece grande importanza alla variabile temporale, poiché quest'ultima ci permette di andare ad analizzare il comportamento degli AS e come questo cambia nel tempo, ad esempio in seguito a mutamenti del mercato o a cambiamenti di atteggiamento da parte dei gestori delle connessioni di determinate reti.

6.2 Primo livello di analisi: la global view

Come detto sopra la global view rappresenta il primo livello di analisi. Si articola su tre differenti visualizzazioni (una bubble chart che visualizza i singoli AS in re-



▲ **tav 01 | Grafico della struttura di BURN**

lazione alle loro caratteristiche di malevolenza, una mappa geografica strutturata su due livelli di zoom, ed un grafico dei trend suddivisi anno per anno), ognuna delle quali serve a mettere in evidenza un aspetto specifico, per offrire una visione d'insieme del fenomeno. Nelle prime due visualizzazioni, la bubble chart e la mappa geografica, la timeline permette di selezionare il periodo di tempo che si desidera analizzare, mentre gli AS, rappresentati come piccole bolle animate, sfruttando diverse variabili visive, ci forniscono svariate informazioni relative all'autonomous system che rappresentano. Nella terza visualizzazione, il grafico dei trend, l'utente può visualizzare l'andamento delle attività malevole anno per anno, individuando velocemente eventuali picchi di attività o comportamenti anomali. Il passaggio da una visualizzazione all'altra è garantito da tre icone poste nell'header della schermata, accanto al logo dell'applicazione.

Prima di andare a presentare le varie visualizzazioni che compongono questo primo livello di analisi si rende necessario andare a delineare quali caratteristiche vengono rappresentate da ogni singolo AS e come. Come detto sopra gli AS sono visualizzati sotto forma di piccole bolle animate, che cambiano in continuazione la loro forma, come fossero delle piccole amebe. Ogni bolla può contenere in se fino a cinque diverse informazioni di tipo visivo:

numero di malicious servers presenti nell'AS: è rappresentato dalla dimensione (area) della bolla rossa che rappresenta l'AS;

dimensione dello spazio IP dell'AS (cioè quanti server possono essere presenti in quella rete): è rappresentato da un cerchio grigio che compare sotto all'AS nel momento in cui lo si seleziona (verrà poi spiegato in dettaglio nei paragrafi a venire);

grado di maliciousness dell'AS: come abbiamo detto ogni AS viene visualizzato sotto forma di una piccola bolla animata, questa animazione varia in base al grado di maliciousness registrato nell'AS. Più un rete è malevola più il movimento della bolla animata sarà intenso e frenetico, a rappresentare una maggiore attività; meno la rete è malevola, meno il movimento della bolla animata sarà intenso e frenetico, fino ad una quasi totale assenza di movimento nelle reti meno malevole;

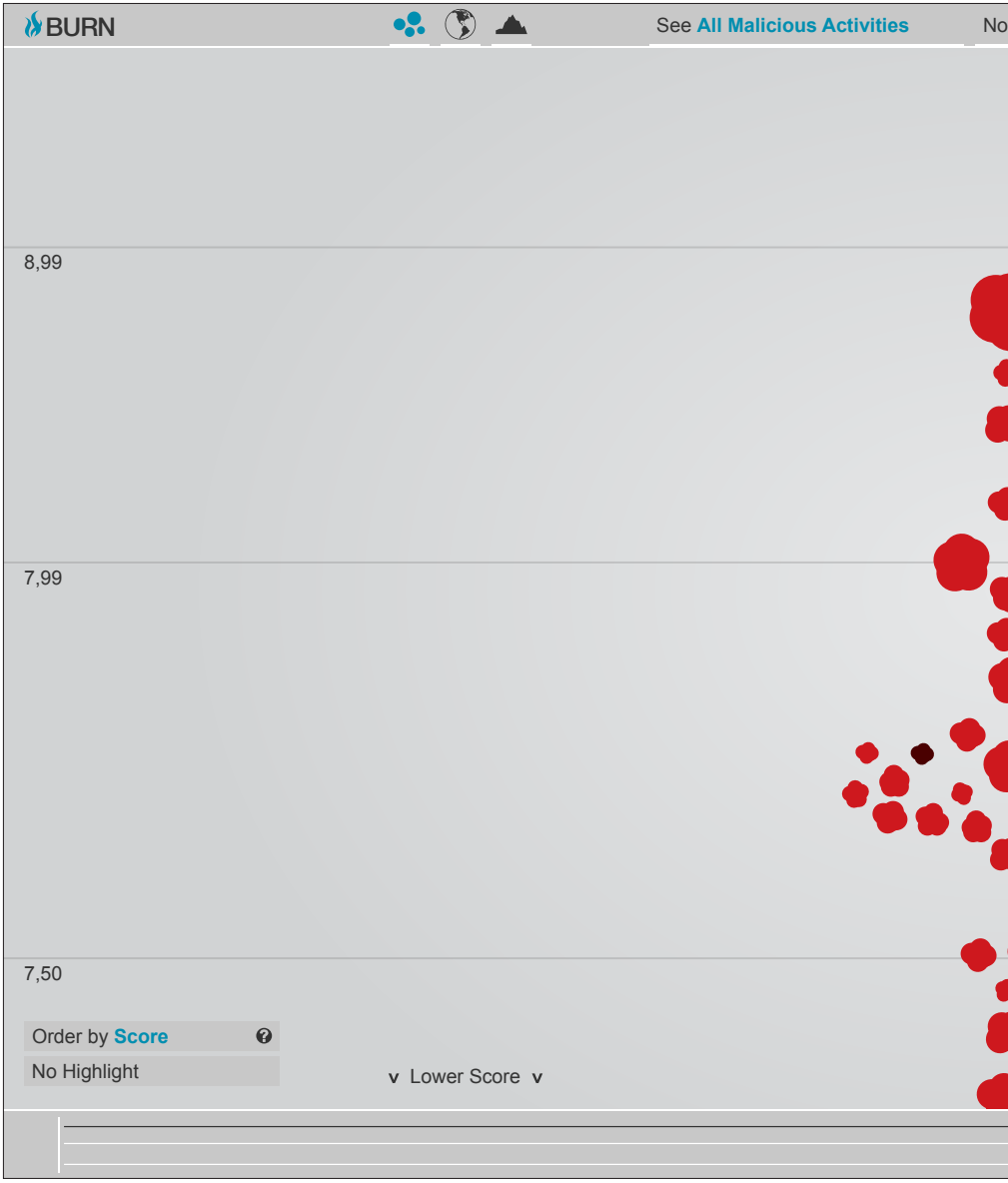
tolleranza dell'AS rispetto all'attività malevola: un AS dove viene registrata attività malevola continuativa per più di un certo periodo viene marcato come tollerante. Agendo sul colore della bolla animata, scurendolo, si facilita all'utente l'individuazione degli AS tolleranti in mezzo agli altri rappresentati;

AS in shutdown: quegli AS dove viene registrato un improvviso calo delle attività malevole, o addirittura una loro totale cessazione, in un periodo breve di tempo, vengono classificati come AS in shutdown, e cioè reti dove l'attività malevola è improvvisamente diminuita o cessata, per spostamento della stessa o per provvedimenti presi contro questo genere di attività. Agendo sulla dimensione e sull'opacità della bolla animata si facilita all'utente l'individuazione degli AS in shutdown in mezzo agli altri rappresentati.

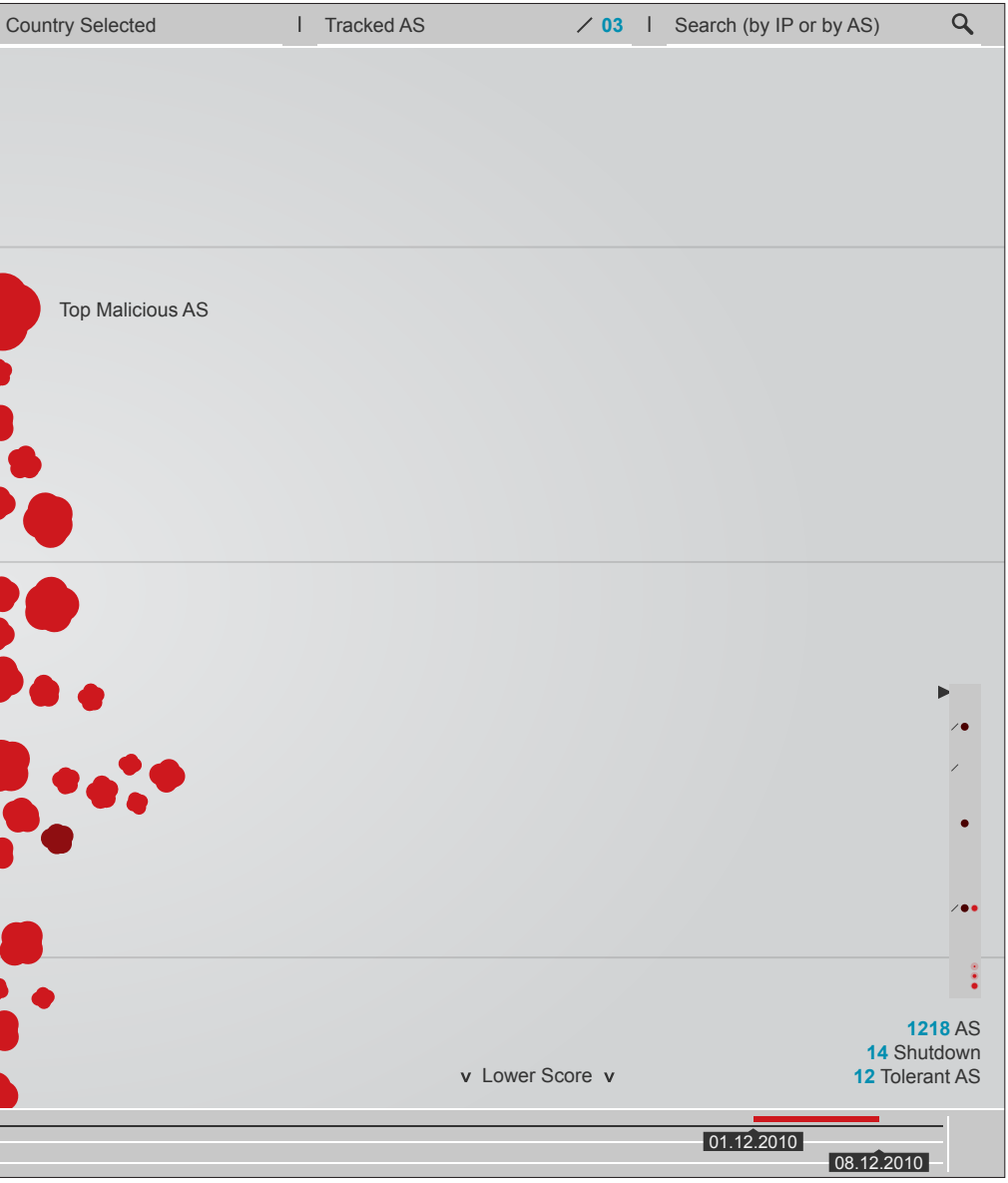
6.2.1 Bubble chart

La bubble chart è la visualizzazione principale di tutto il sistema, la prima dove l'utente atterra nel momento in cui accede a BURN. Si compone di una mappa dove gli AS, rappresentati come sopra descritto come piccole bolle animate, sono posizionati in base al loro livello di malicious score (il livello di malevolenza legato ad ogni AS dal database di FIRE, che va a comporre giorno per giorno una classifica degli AS stessi). Questa variabile, posizionata sull'asse verticale, permette di ordinare gli AS, partendo dall'alto con quello col più alto grado di malicious score registrato, per poi discendere. Volutamente non è stato assegnato nessun valore all'asse delle x: la posizione orizzontale degli AS viene automaticamente calcolata in maniera tale da ottimizzare lo spazio tra le bolle ed evitare sovrapposizioni. Lasciando la dimensione orizzontale libera da vincoli, gli AS si dispongono unicamente in base al valore presente sull'asse delle y, creando degli agglomerati, dei cluster visivi, in prossimità di valori simili sulla dimensione verticale.
Questo può risultare utile per raggruppare visivamente, secondo il principio della prossimità analizzato all'inizio del secolo dalla Gestalt School of Psychology, quegli AS che sono accomunati da caratteristiche simili. L'intera visualizzazione può essere scrollata verso il basso, per andare ad indagare tutti gli AS registrati nel periodo preso in considerazione, mentre sullo sfondo sono sempre presenti dei riferimenti ad una scala di valori, che aiuta a mantenere sott'occhio la situazione.

▶ tav 02 | La Bubble chart



6. BURN, Baring Unknown Rogue Networks


Country Selected
Tracked AS
03
Search (by IP or by AS)
Top Malicious AS
1218 AS
14 Shutdown
12 Tolerant AS
v Lower Score v
01.12.2010
08.12.2010

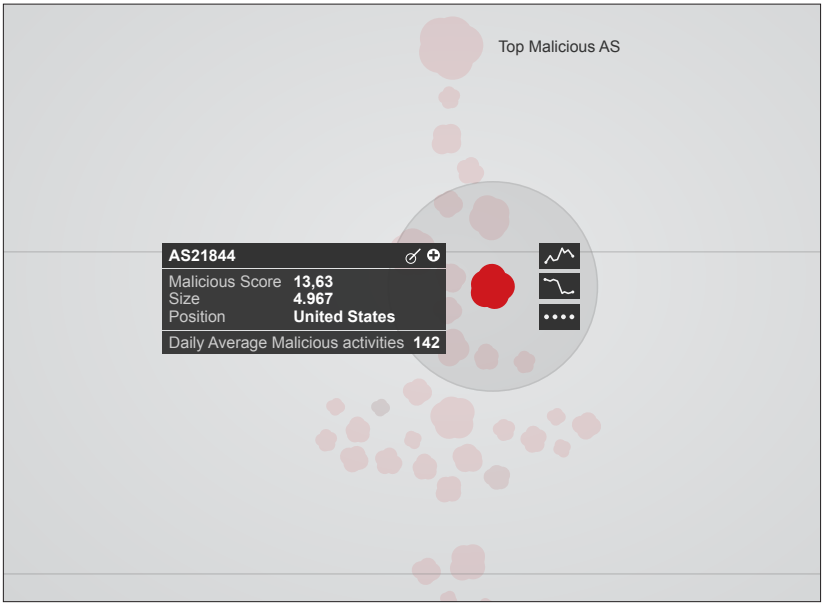
Cliccando sulle bolle si seleziona il singolo AS, questa selezione comporta l'apertura di un box contestuale che contiene alcune informazioni relative all'AS, la visualizzazione della dimensione dello spazio IP dell'AS, e l'apparizione di un menù che permette di andare ad indagare più a fondo alcuni aspetti specifici dell'AS stesso, accedendo al secondo livello di analisi.

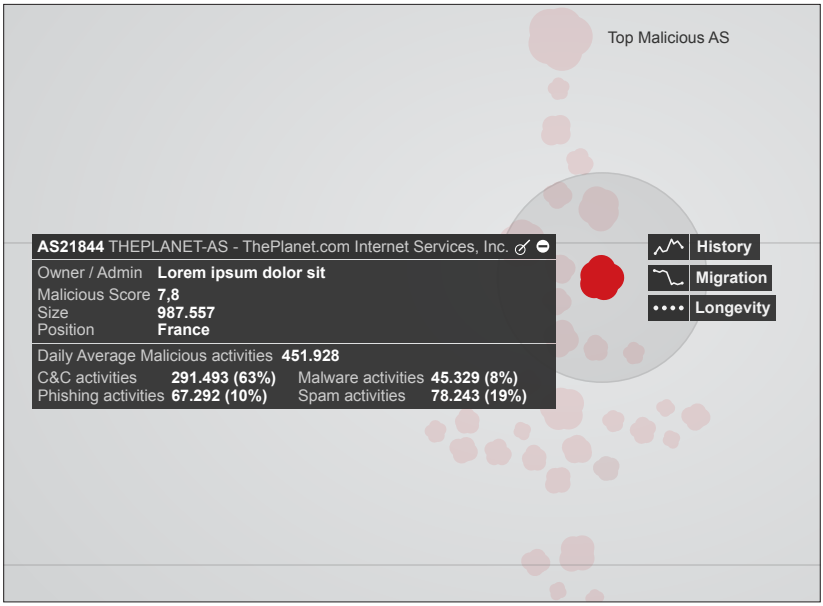
Il box contestuale, che appare sulla sinistra della bolla rappresentante l'AS, si struttura anch'esso su due livelli: un primo più compatto, con le informazioni base, ed uno esteso, con alcune informazioni aggiuntive. Nella sua forma compatta, quella che si apre di default cliccando sulla bolla animata, il box contiene il numero identificativo dell'AS, il suo valore di malicious score, la sua size, lo stato di appartenenza ed il numero medio di attività malevole registrate nel periodo selezionato. Nella forma estesa queste informazioni vengono integrate con il nome commerciale esteso dell'AS, il nome dell'amministratore della rete, e una suddivisione in percentuale di come l'attività malevola registrata sull'AS si suddivide tra le quattro categorie analizzate (C&C, malware, phishing, spam). Dal box contestuale è possibile (cliccando sull'icona presente nell'angolo in alto a destra dello stesso, posizionata accanto all'icona per espandere o contrarre il box) aggiungere l'AS alla autonomous system tracking list (le cui funzionalità verranno spiegate nello specifico nel capitolo 6.4.4 "Autonomous system tracking list") per poterne mantenere traccia. Inoltre, per facilitare la comprensione di tutti gli elementi presenti ad un pubblico anche non esperto, passando in mouse over sui vari dati presenti nel box questi vengono evidenziati sulla visualizzazione dell'AS.

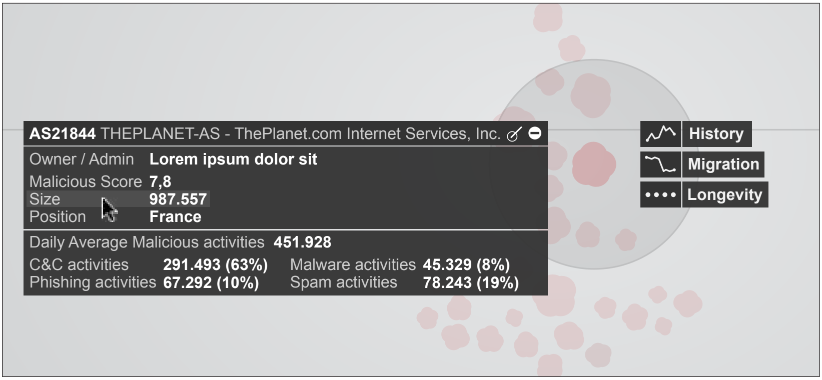
Sulla destra dell'AS appare invece un menù composto da tre bottoni che permettono l'accesso rispettivamente a history chart, service migration screen e longevity analysis chart. Queste sezioni verranno in seguito spiegate nel capitolo relativo al secondo livello di analisi: l'autonomous system view.

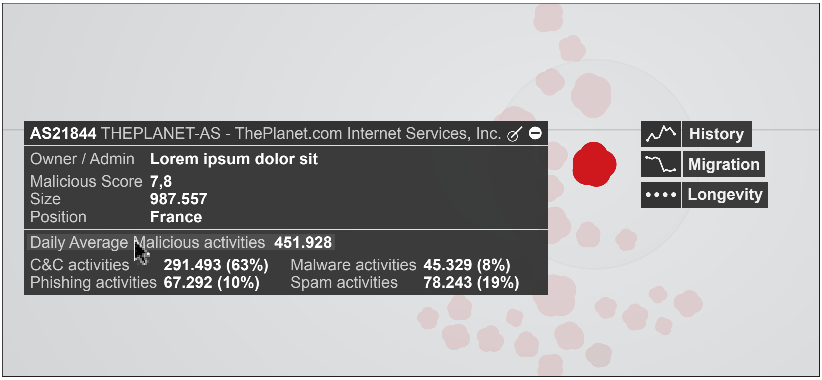
▶ tav 03a-b | Il box contestuale

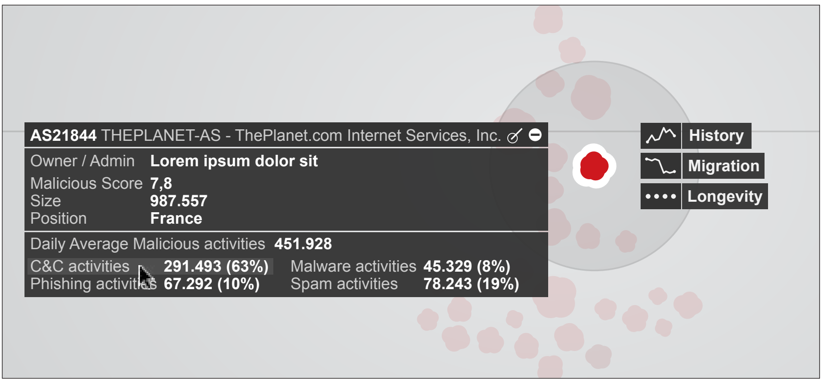
▶▶ tav 04a-b-c | Esempi di mouse over sul box contestuale


Top Malicious AS
AS21844
Malicious Score 13,63
Size 4.967
Position United States
Daily Average Malicious activities 142


Top Malicious AS
AS21844 THEPLANET-AS - ThePlanet.com Internet Services, Inc.
Owner / Admin Lorem ipsum dolor sit
Malicious Score 7,8
Size 987.557
Position France
Daily Average Malicious activities 451.928
C&C activities 291.493 (63%)
Malware activities 45.329 (8%)
Phishing activities 67.292 (10%)
Spam activities 78.243 (19%)
History
Migration
Longevity

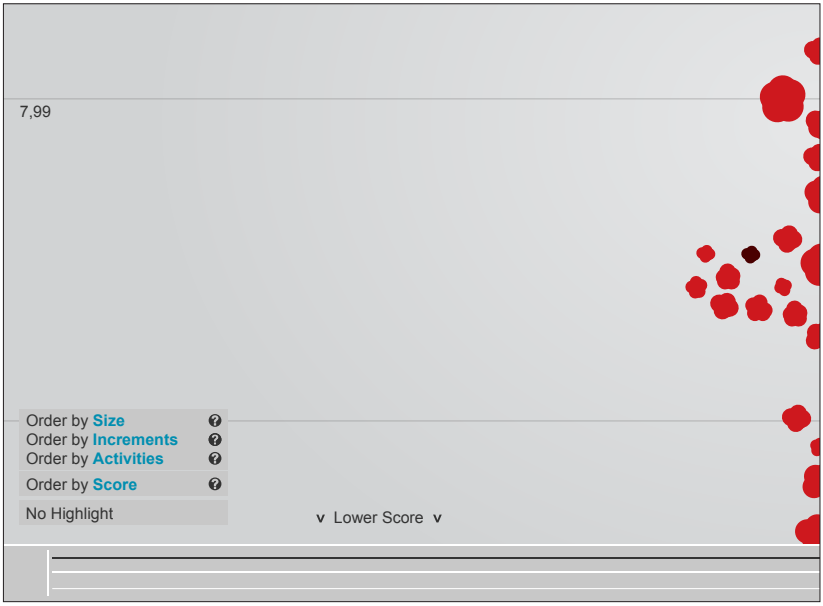

AS21844 THEPLANET-AS - ThePlanet.com Internet Services, Inc.
Owner / Admin Lorem ipsum dolor sit
Malicious Score 7,8
Size 987.557
Position France
Daily Average Malicious activities 451.928
C&C activities 291.493 (63%)
Malware activities 45.329 (8%)
Phishing activities 67.292 (10%)
Spam activities 78.243 (19%)
History
Migration
Longevity

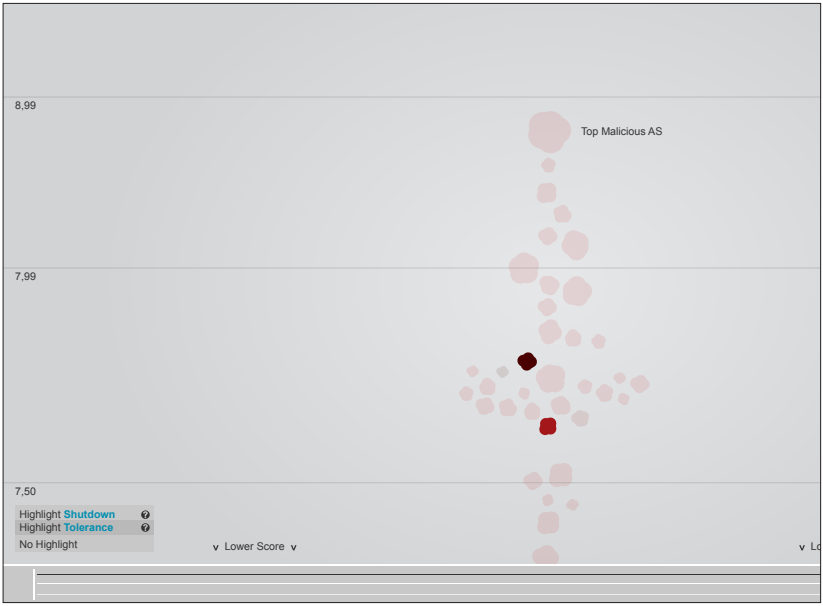

AS21844 THEPLANET-AS - ThePlanet.com Internet Services, Inc.
Owner / Admin Lorem ipsum dolor sit
Malicious Score 7,8
Size 987.557
Position France
Daily Average Malicious activities 451.928
C&C activities 291.493 (63%)
Malware activities 45.329 (8%)
Phishing activities 67.292 (10%)
Spam activities 78.243 (19%)
History
Migration
Longevity


AS21844 THEPLANET-AS - ThePlanet.com Internet Services, Inc.
Owner / Admin Lorem ipsum dolor sit
Malicious Score 7,8
Size 987.557
Position France
Daily Average Malicious activities 451.928
C&C activities 291.493 (63%)
Malware activities 45.329 (8%)
Phishing activities 67.292 (10%)
Spam activities 78.243 (19%)
History
Migration
Longevity

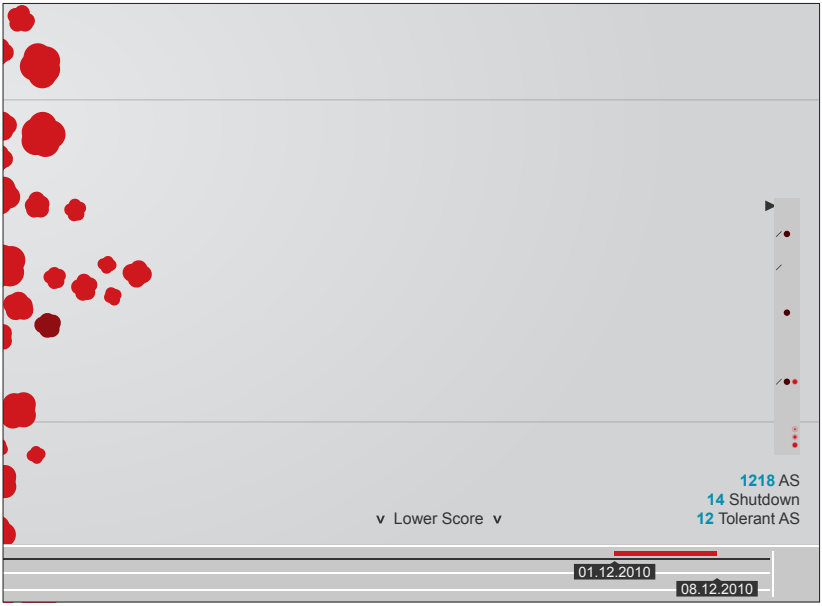
Nell'angolo inferiore sinistro di questa visualizzazione sono presenti due strumenti utili per personalizzare la mappa che si sta osservando:
- un menù a tendina permette di cambiare il valore relativo all'asse delle y. L'intera visualizzazione difatti, ordinata di default per malicious score, può essere ordinata anche per AS size (la dimensione dello spazio IP di ogni singolo AS), per increments (variazione calcolata giorno per giorno dell'attività malevola presente su un AS), o per activities (la tipologia di attività scelta attraverso l'activity filter, che verrà spiegato in seguito). Al momento sono stati inseriti in questo menù soltanto tre valori, ma in futuro questi potrebbero anche aumentare, per analizzare altri comportamenti relativi alle reti. Il livello di malicious score di ogni AS viene in ogni caso sempre visivamente mantenuto dall'animazione presente sulla bolla (come spiegheremo più in dettaglio nei capitoli a venire);
- un filtro per evidenziare gli AS in shutdown o gli AS registrati come tolleranti. Questo filtro, combinato con le già presenti animazioni, aiuta ulteriormente l'utente ad individuare nella massa quegli AS che presentano comportamenti specifici.

▶ **tav 05 | Menù per ordinare l'asse y**
▶▶ **tav 06 | Filtro Highlight**





6. BURN, Baring Unknown Rogue Networks



Nell'angolo inferiore destro della visualizzazione un semplice navigatore aiuta a mantenere sott'occhio la posizione dell'utente all'interno della mappa, e ad evidenziare le zone dove sono presenti AS con comportamenti interessanti, mentre alcuni contatori riportano il numero totale di AS rappresentati nella visualizzazione, quanti di essi sono in shutdown e quanti sono classificati come tolleranti.

▲ tav 07 | Navigatore e contatori

6.2.2 Geographical map

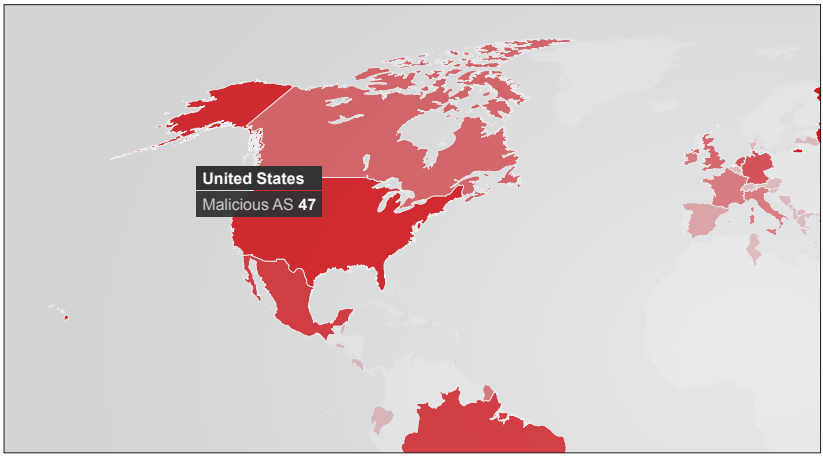
La geographical map è una visualizzazione geografica dove gli AS vengono posizionati in base alle coordinate di registrazione della rete. Strutturata su due livelli di zoom, mondiale e singolo stato, permette all'utente di vedere la distribuzione fisica delle reti malevole.

La mappa a livello mondiale è costituita da una *single-hue choropleth map* [1], dove i singoli stati sono colorati di un rosso più o meno intenso a seconda di un livello di malicious score calcolato prendendo il numero di AS malevoli individuati in ogni stato e normalizzandolo per il numero totale di AS presenti nello stesso (per ovviare alle problematiche presentate in EMBER, capitolo 3.1.2), più alto il livello di malicious score, più intensa la saturazione del colore dello stato. Questa visualizzazione fornisce un rapido colpo d'occhio su quali sono le regioni del mondo maggiormente colpite da questo fenomeno, quali sono gli stati maggiormente implicati e così via. Inoltre bisogna ricordare che in stati diversi vigono leggi diverse, più o meno improntate ad una risoluzione o perlomeno ad un contenimento del fenomeno, e la distribuzione a livello mondiale dell'attività malevola può anche essere conseguenza di diverse politiche o addirittura di specifici avvenimenti storici (vedi gli esempi fatti nel capitolo relativo al cyber crime ed alla cyber warfare). Passando col mouse sui singoli stati si aprono dei pop-up che visualizzano il nome dello stato ed il numero di malicious AS individuati al suo interno.

In questa visualizzazione nell'angolo inferiore sinistro è presente un menù a tendina che in questo caso permette di cambiare il valore secondo il quale gli stati vengono colorati. Di default colorati per malicious score, possono essere colorati anche per numero di shutdown registrati nello stato o numero di AS considerati tolleranti presenti nello stato. Nell'angolo inferiore destro sono sempre presente i contatori che riportano il numero totale di AS presenti a livello mondiale, quanti di essi sono in shutdown e quanti sono classificati come tolleranti.

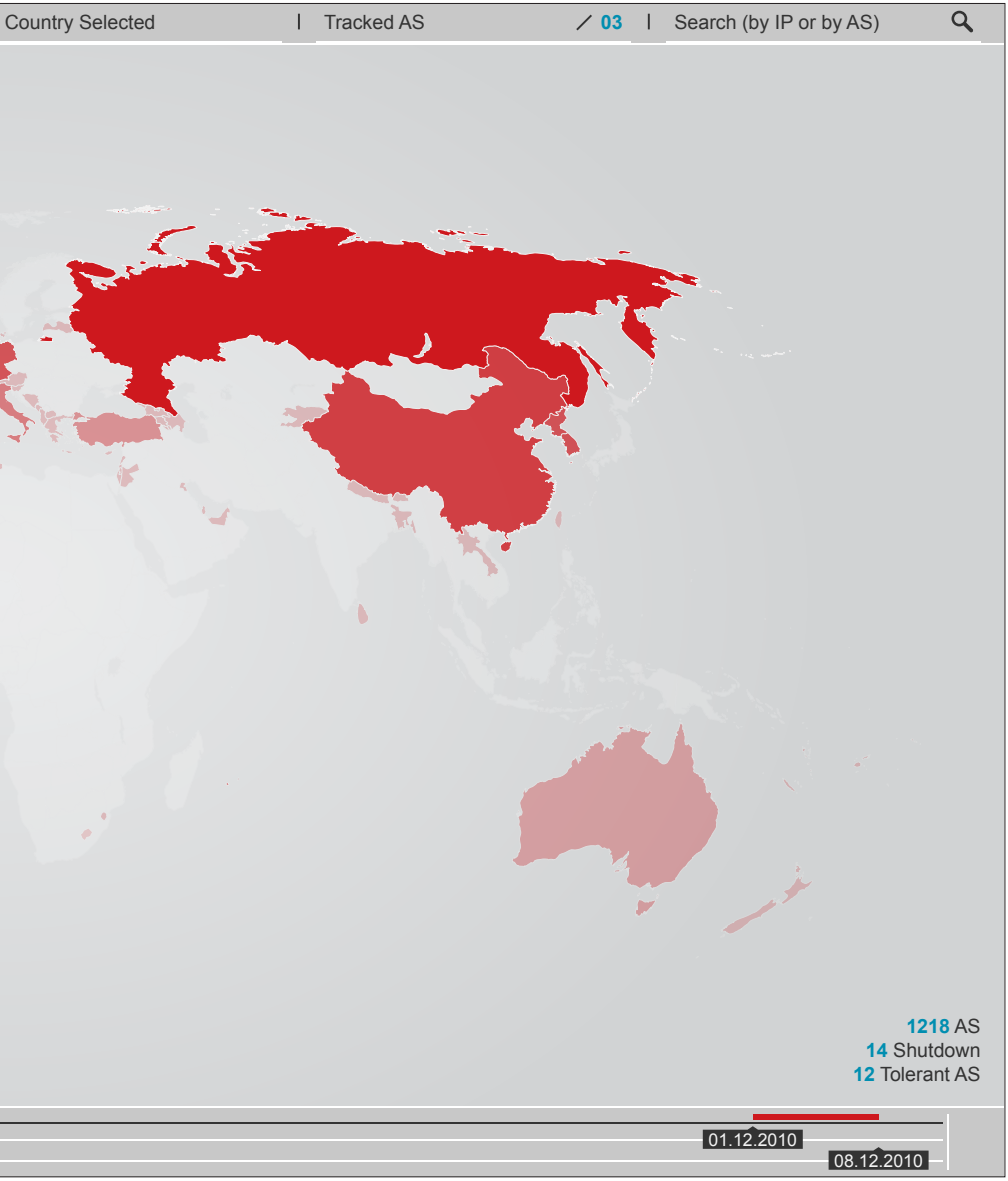
[1] A. H. Robinson, J. L. Morrison, P. C. Muehrcke, A. J. Kimerling, and S. C. Guptill. Elements of Cartography. Wiley, 6 edition, Mar. 1995.

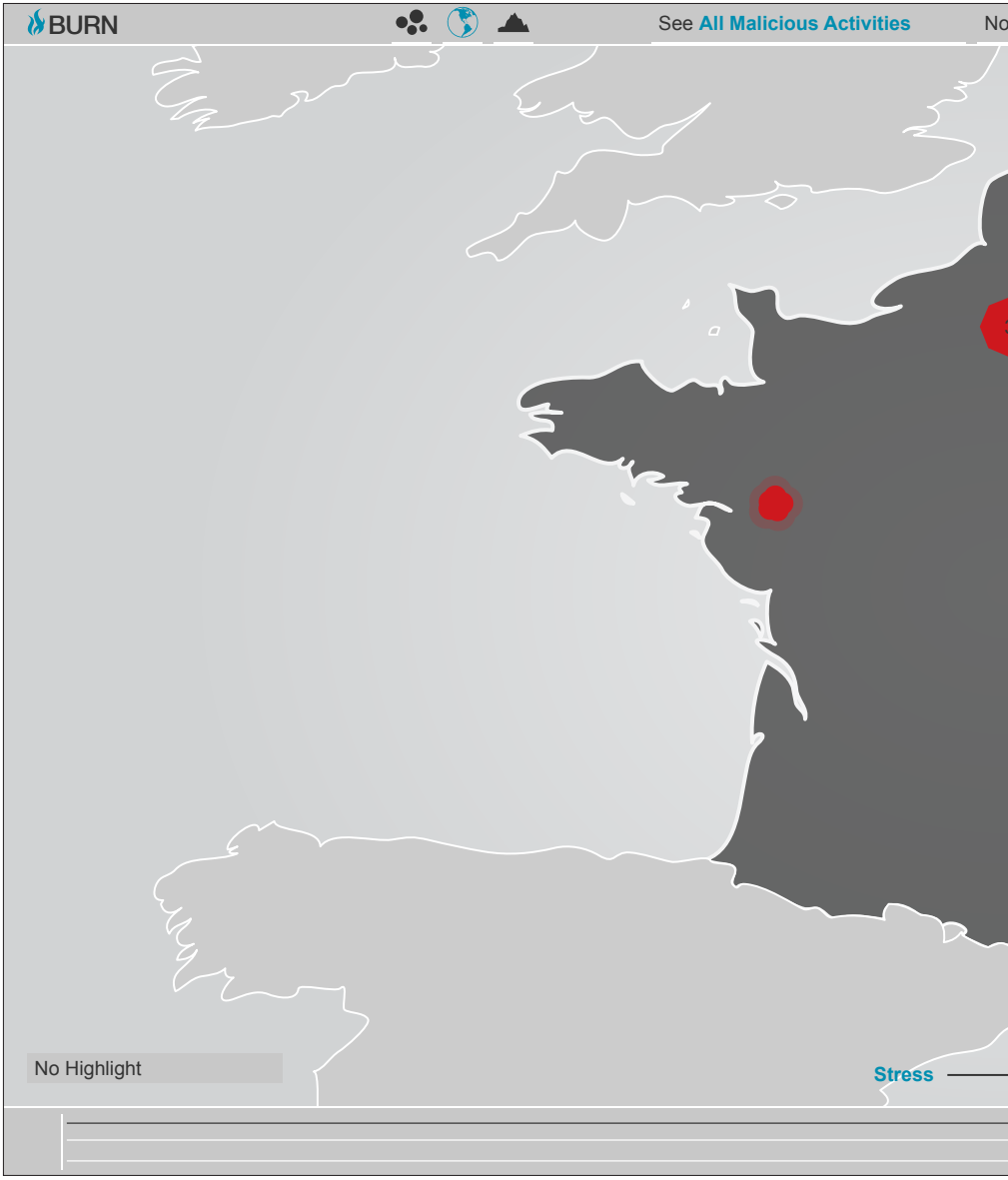
Dalla mappa a livello mondiale, l'utente può cliccare sui singoli stati per zoommare su di essi. In questo modo si passa alla visualizzazione del singolo stato dove gli AS, nuovamente rappresentati sotto forma di piccole bolle animate, con le stesse caratteristiche precedentemente spiegate, sono posizionati su di una mappa tematica in base alle coordinate con cui la rete è stata registrata.
In questa visualizzazione nell'angolo inferiore sinistro è presente il filtro che permette l'evidenziazione degli AS in shutdown o tolleranti (lo stesso filtro che era già presente nella Bubble chart), mentre nell'angolo inferiore destro vengono mantenuti i contatori, questa volta relativi ai dati del singolo stato visualizzato. In aggiunta a questo è qui presente anche uno strumento chiamato stress che serve per aumentare o diminuire in percentuale la dimensione di tutte le bolle animate, in maniera tale da poter ottimizzare la visualizzazione nel caso di una presenza massiccia di AS.
Un bottone posizionato nell'angolo in alto a destra permette di tornare al livello di zoom mondiale.



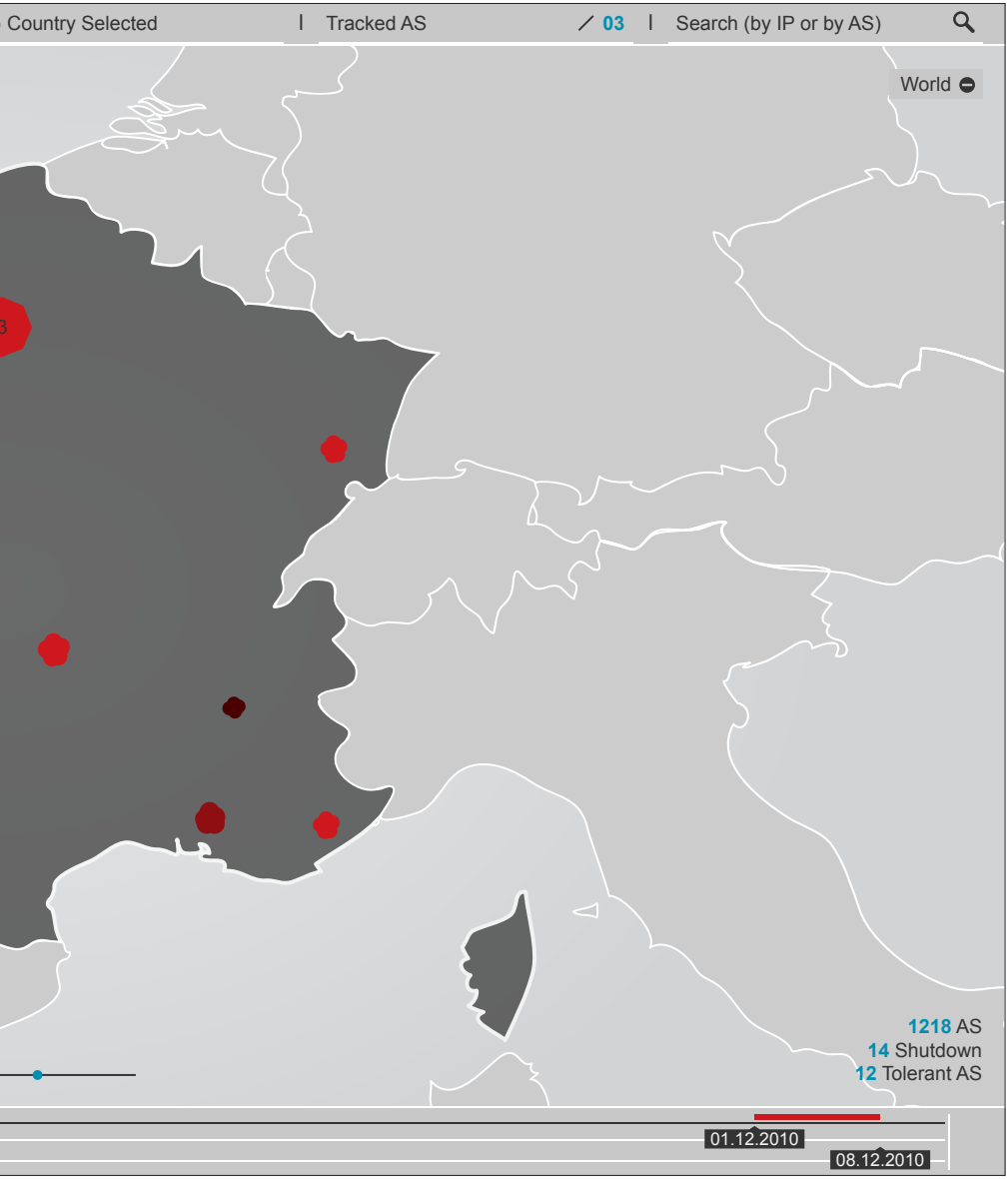
▲ tav 08 | Pop-up contestuale sul singolo stato
▶ tav 09 | La geographical map a livello mondiale
▶▶ tav 10 | La geographical map zoommata sul singolo stato

6. BURN, Baring Unknown Rogue Networks

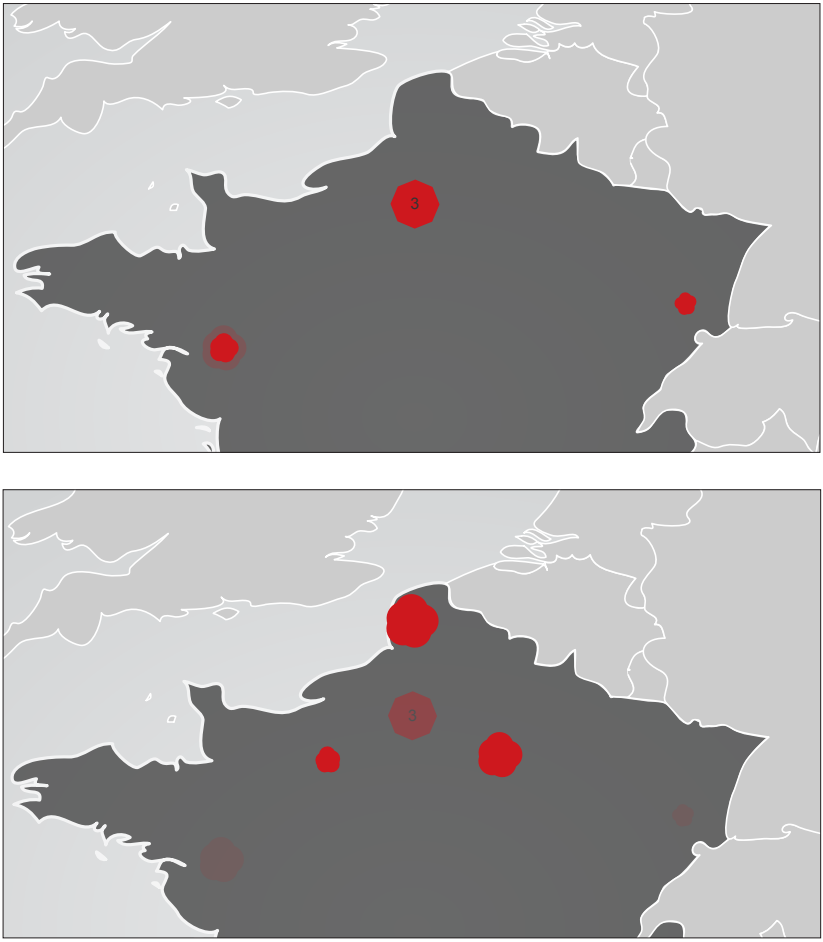

Country Selected
Tracked AS
03
Search (by IP or by AS)
1218 AS
14 Shutdown
12 Tolerant AS
01.12.2010
08.12.2010



6. BURN, Baring Unknown Rogue Networks


Country Selected
Tracked AS
03
Search (by IP or by AS)
World
1218 AS
14 Shutdown
12 Tolerant AS
01.12.2010
08.12.2010

In questa visualizzazione sono presenti anche degli elementi, di forma ottagonale, differenti dalle bolle animate, sempre posizionati sulla mappa. Poiché alcuni AS vengono registrati alle stesse coordinate, si è resa necessaria la creazione di questi collettori di AS che contengono al loro interno più reti. Cliccando su di essi è possibile espanderli per osservare tutti gli AS registrati in quelle coordinate.



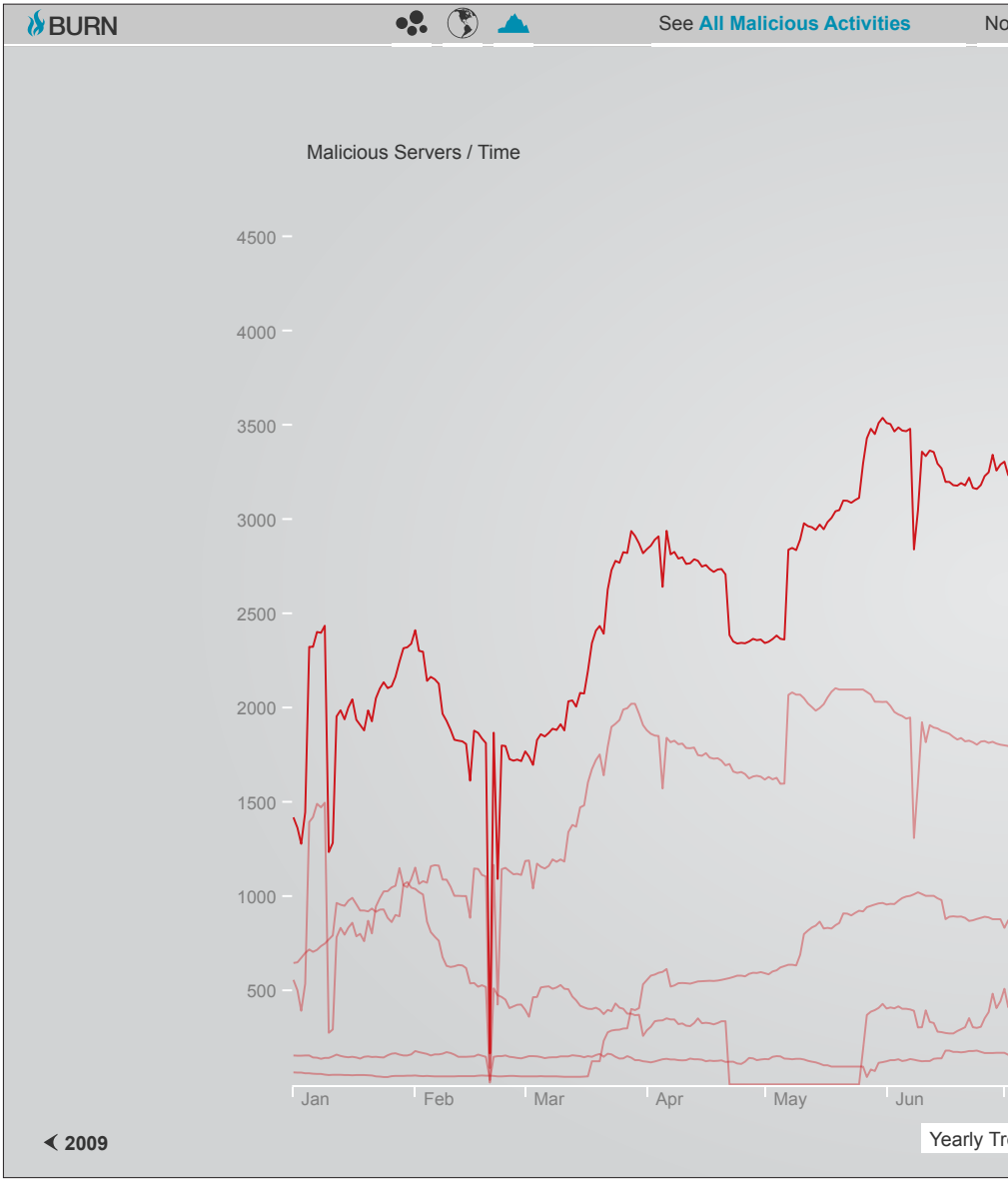
▲ **tav 11a-b | Esempio di collettore, chiuso e aperto**

Come scelta progettuale si è deciso di utilizzare due livelli di zoom fissi, e non uno zoom dinamico, mantenendo lo zoom maggiore, quello sul singolo stato, ad una risoluzione abbastanza bassa. Questa scelta è stata presa principalmente per due motivi:
- la geolocalizzazione delle aree coperte da ogni singola rete è per sua stessa natura imprecisa, tutto quello che i dati possono fornire sono le coordinate di registrazione commerciale della rete e la dimensione della stessa, di sicuro non il posizionamento preciso di tutti i singoli server presenti nella rete;
- offrire un livello di zoom maggiore di quello dato potrebbe risultare ingannevole per gli utenti meno esperti, portandoli, ad esempio, a pensare di poter localizzare all'interno di una città il punto esatto in cui la rete segnalata come malevola opera.
Per l'analisi di questo fenomeno i livelli di zoom proposti sono sufficienti per coprire tutte le informazioni che attualmente possono essere registrate, analizzate, e, di conseguenza, visualizzate.

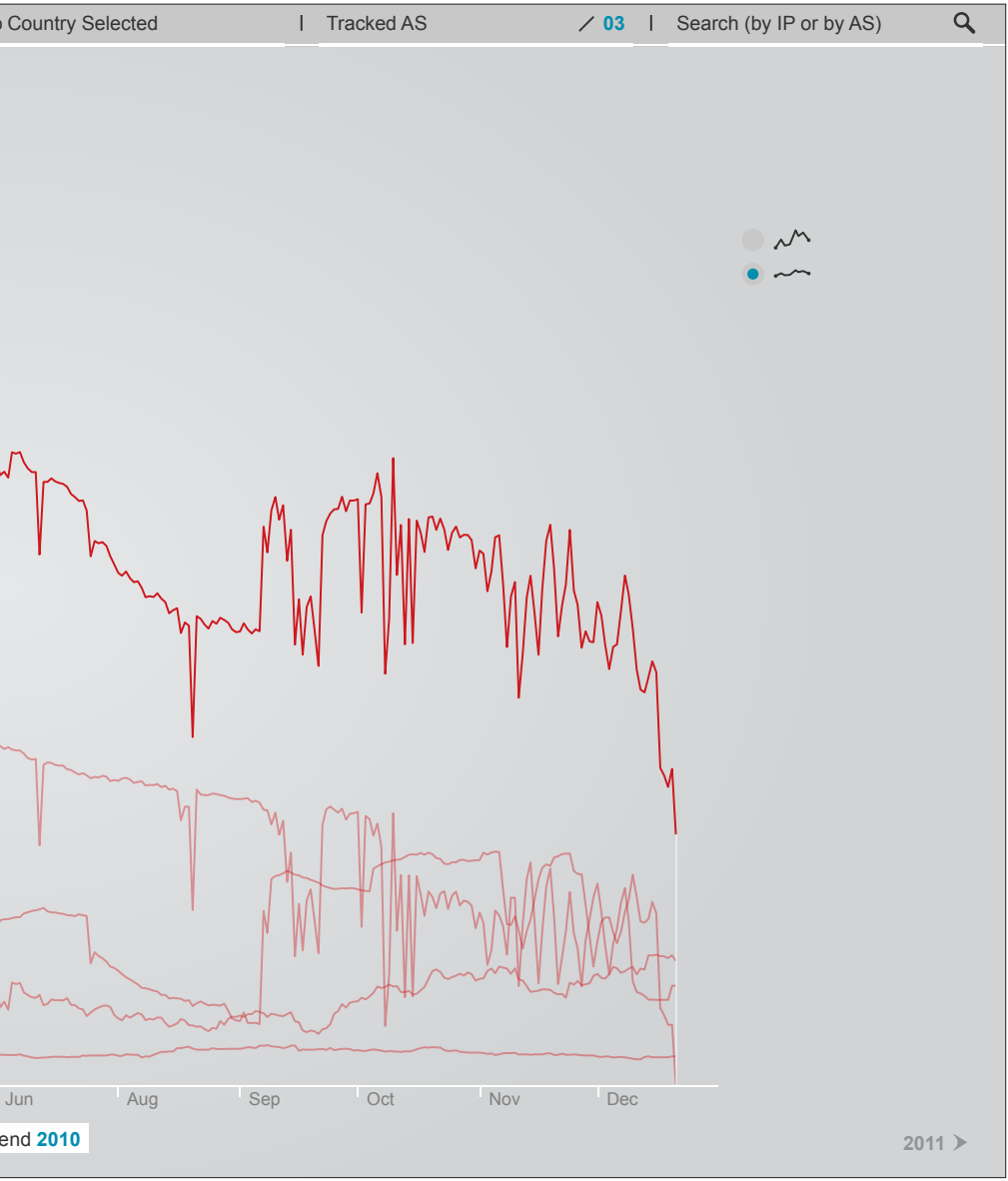
6.2.3 Trend chart

Nella trend chart, ultima delle tre visualizzazioni appartenenti alla global view, il numero dei malicious servers, registrati a livello globale, viene messo in relazione al tempo, attraverso l'utilizzo di un grafico multi-line plot. In questa visualizzazione è presente una linea per ogni tipologia di attività (c&c, malware, phishing, spam) più una linea per la somma di tutte le attività. L'asse temporale è suddiviso anno per anno e l'utente, attraverso due bottoni posti ai lati del grafico, può cambiare l'anno che si sta visualizzando. Questa tipologia di visualizzazione è utile per osservare gli andamenti a livello globale e individuare picchi o interruzioni dell'attività malevola, permette di avere uno spaccato sulla quantità di attività malevola registrata nel corso di un determinato anno, e, suddividendola tra le diverse tipologie, di farsi un'idea su quali siano state le tipologie di attività maggiormente utilizzate.

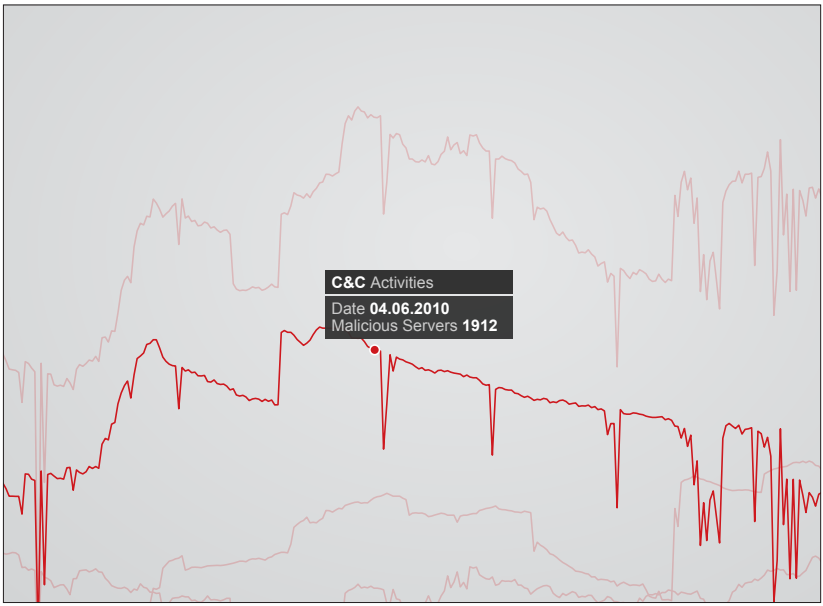
▶ tav 12 | La trend chart



6. BURN, Baring Unknown Rogue Networks


Country Selected
Tracked AS
03
Search (by IP or by AS)
Jun
Aug
Sep
Oct
Nov
Dec
end 2010
2011

Passando il mouse sulle diverse linee che compongono il grafico le si portano in evidenza rispetto alle altre e si aprono dei piccoli pop-up che, oltre a dirci a quale tipologia di attività appartiene la linea su cui ci siamo posizionati, forniscono indicazioni più precise su giorno e numero di malicious servers relativi al punto in cui l'utente ha posto il puntatore.
Inoltre sulla destra del grafico è presente un menù a radio button che permette di scegliere se forzare o meno la visualizzazione ad ottimizzare tutto lo spazio disponibile. Questa opzione è utile in quanto quando in questa visualizzazione si seleziona un singolo stato il grafico si riduce notevolemente di dimensioni, fornendo un rapporto corretto rispetto al totale, ma renderndo di difficile lettura le linee del grafico. Con questo controller è possibile espandere la dimensione verticale per sfruttare tutto lo spazio a disposizione e permettere una più agevole lettura dei dati.



▲ **tav 13 | Pop-up trend chart**

▶ **tav 14a-b | Esempio di funzionamento dei radio button nella trend chart**


BURN
See All Malicious Activities
Australia
Tracked AS
03
Search (by IP or by AS)
Malicious Servers / Time
4500
4000
3500
3000
2500
2000
1500
1000
500
Jan
Feb
Mar
Apr
May
Jun
Jun
Aug
Sep
Oct
Nov
Dec
2009
Yearly Trend 2010
2011


BURN
See All Malicious Activities
Australia
Tracked AS
03
Search (by IP or by AS)
Malicious Servers / Time
450
400
350
300
250
200
150
100
50
Jan
Feb
Mar
Apr
May
Jun
Jun
Aug
Sep
Oct
Nov
Dec
2009
Yearly Trend 2010
2011

6.3 Secondo livello di analisi: l'autonomous system view

Il secondo livello di analisi, l'autonomous system view, sposta l'attenzione dal globale al singolo, andando ad analizzare più nello specifico il comportamento dei singoli AS con l'ausilio di tre ulteriori visualizzazioni.
Come descritto in precedenza (tav 03a-b) cliccando sulle bolle animate, presenti nella bubble chart o nella geographical map zoommata sul singolo stato, si può selezionare il singolo AS aprendone un box contestuale, sulla sinistra, ed un menù composto da tre bottoni, sulla destra. Attraverso questo menù l'utente può accedere al secondo livello di analisi, incentrato sul singolo AS, scegliendo tra tre sezioni disponibili: history chart, service migration screen e longevity analysis chart. La history chart contiene dei grafici relativi agli andamenti del singolo AS nel tempo, la service migration screen serve ad individuare possibili migrazioni di attività, la longevity analysis chart serve ad analizzare il livello di tolleranza dell'AS nei confronti di quei server che perpetrano attività malevola in maniera continuativa. Tre icone, poste nella parte alta al centro della visualizzazione, permettono il passaggio da una sezione all'altra.

In tutte le tre tipologie di visualizzazione appartenenti al secondo livello di analisi vi sono due elementi ricorrenti:
- nell'angolo in alto a sinistra un'etichetta fornisce numero e nome commerciale dell'AS di cui si stanno visualizzando i dettagli. Similmente ai box contestuali descritti nel capitolo relativo alla bubble chart, queste etichette possono essere espanse per visionare le informazioni relative all'AS, ed anche qui è presente l'icona che permette di aggiungere, o rimuovere, l'AS dalla autonomous system tracking list;
- nell'angolo in alto a destra un bottone permette di chiudere la visualizzazione corrente e tornare così alla global view.

A monte della realizzazione di due delle tre visualizzazioni (service migration screen e longevity analysis chart) presenti in questo secondo livello di analisi, ed anche per alcune parti delle visualizzazioni presentate nella global view, vi è stato lo studio e la creazione di una serie di algoritmi che, sfruttando i dati presenti nel database di FIRE, analizzandoli e combinandoli tra loro, hanno reso possibile la

creazione di nuovi dati ed informazioni, prima non disponibili. Per un maggiore approfondimento su questo fare riferimento alla tesi di Luca Di Mario.

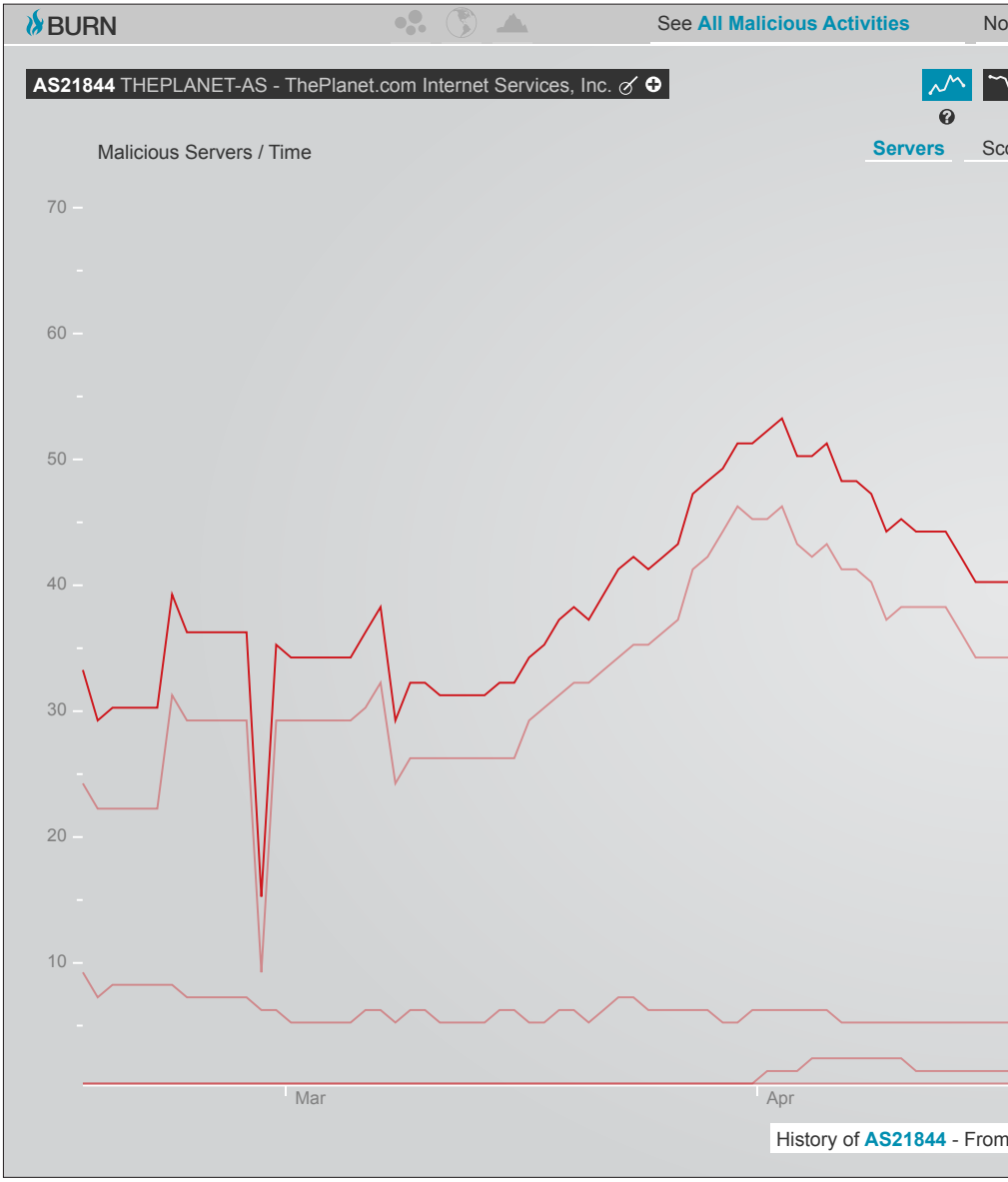
6.3.1 History chart

La history chart è una trend chart (vedi capitolo sulla global view) ridotta a quattro mesi di attività (quattro mesi calcolati a partire dall'ultimo giorno del periodo selezionato tramite la timeline e spostandosi indietro nel tempo) registrata sul singolo AS selezionato. A differenza della trend chart, in questa visualizzazione sono presenti tre bottoni, posizionati in alto al centro, che permettono di scegliere quale variabile visualizzare sull'asse verticale del grafico:
- servers: mostra il numero di malicious servers, come avviene anche nella trend chart;
- score: mostra il valore di malicious score associato all'AS;
- rank: mostra la posizione dell'AS in relazione ad una classifica globale di tutti gli AS (classifica stilata rispetto al livello di malicious score registrato per ogni AS).

Come per la trend chart, anche in questo grafico vengono disegnate più linee, una per ogni tipologia di attività, più una per la somma di tutte le attività. Anche qui passando in mouse over sulle diverse linee le si evidenziano e si aprono i pop-up informativi.

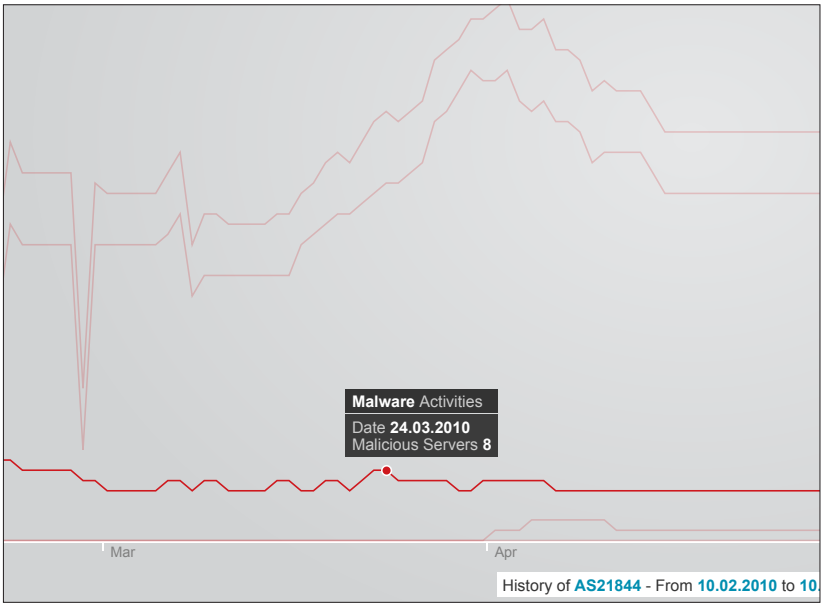
Una fascia sullo sfondo ci ricorda qual'è il periodo di attività che abbiamo deciso di analizzare attraverso la timeline nella global view.

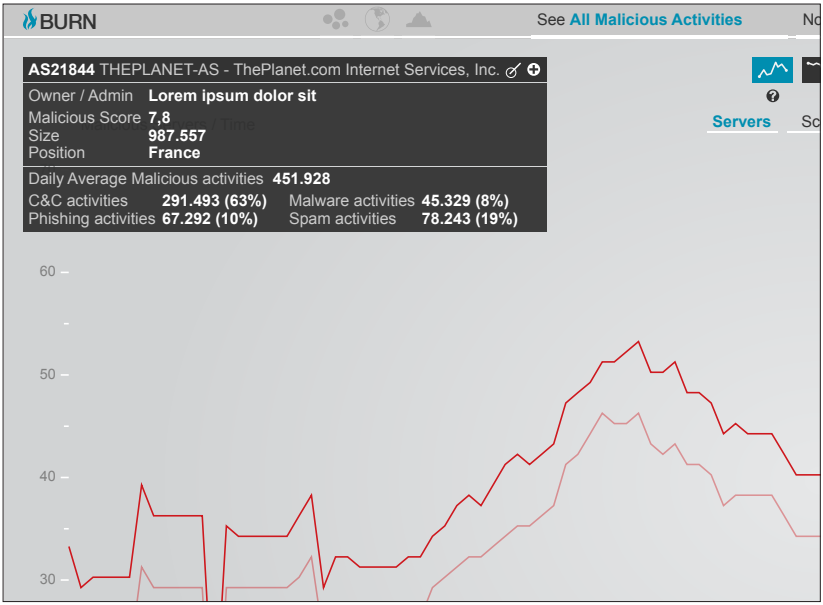
▶ tav 15 | La history chart
▶▶ tav 16 | Pop-up history chart
▶▶▶ tav 17 | Box contestuale espanso all'interno dell'autonomous system view



6. BURN, Baring Unknown Rogue Networks


Country Selected
Tracked AS
03
Search (by IP or by AS)
Close
ore
Rank
May
Jun
10.02.2010 to 10.06.2010





6. BURN, Baring Unknown Rogue Networks

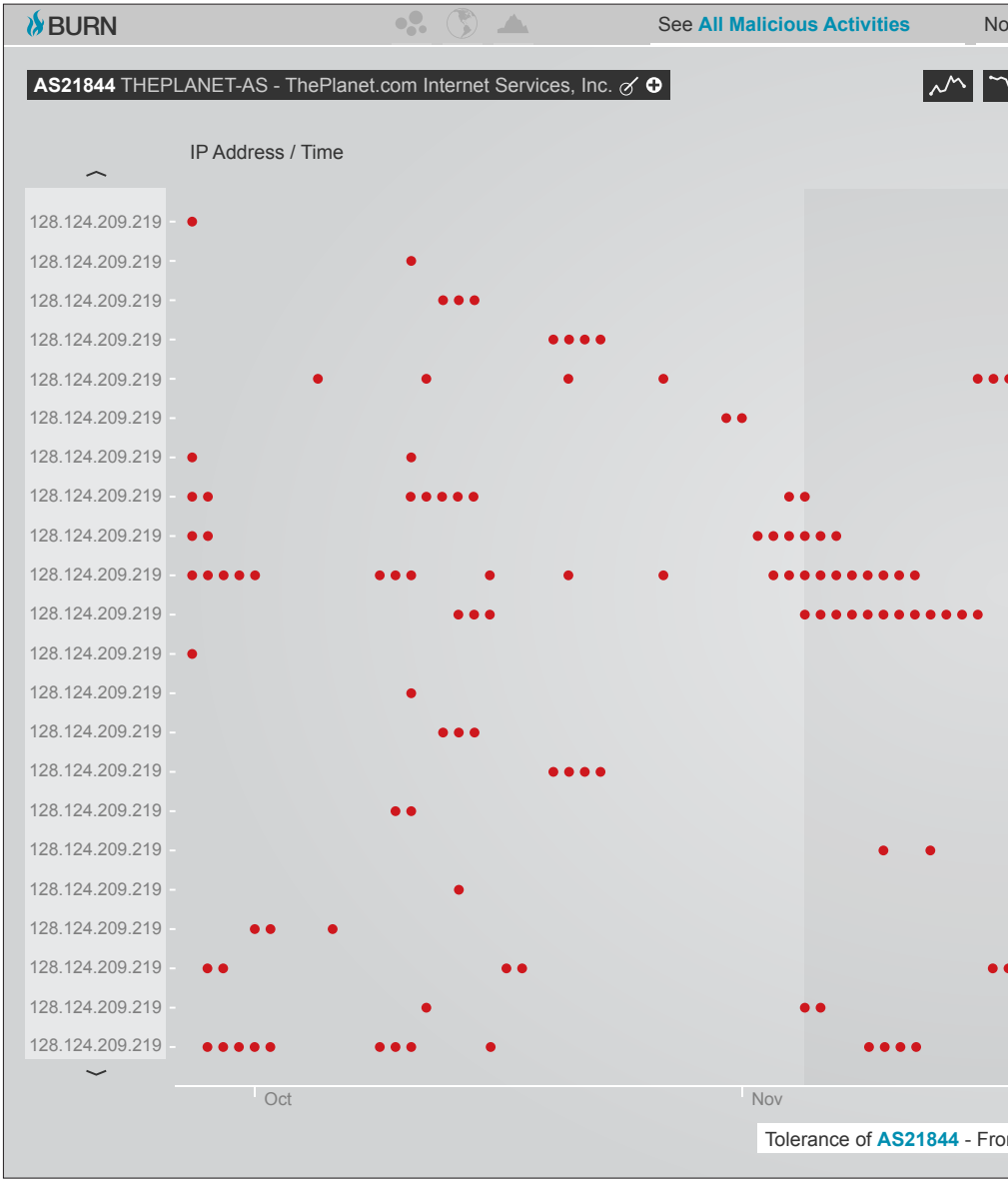
6.3.2 Service longevity chart

Nonostante il database di FIRE, il database sui cui dati si basa BURN, faccia già una selezione, non interessandosi a quegli AS che non dimostrano un persistente e continuativo atteggiamento malevolo, nelle fasi preliminari del progetto abbiamo osservato come alcuni AS presentino importanti periodi di attività interrotti da periodi di intermittenza nella stessa. Questo fenomeno ci ha spinto a creare una sezione, la Service longevity chart, che, anche in questo caso, fosse di supporto all'analisi manuale del fenomeno. In questa sezione BURN permette all'utente di visualizzare giorno per giorno la presenza o meno di attività malevola all'interno dei singoli server presenti in un AS, stabilendo, in base alla quantità di server attivi e da quanto tempo agiscono in maniera indisturbata, se quell'AS è da considerarsi tollerante, e segnalandolo in questo modo come tale anche nella global view.

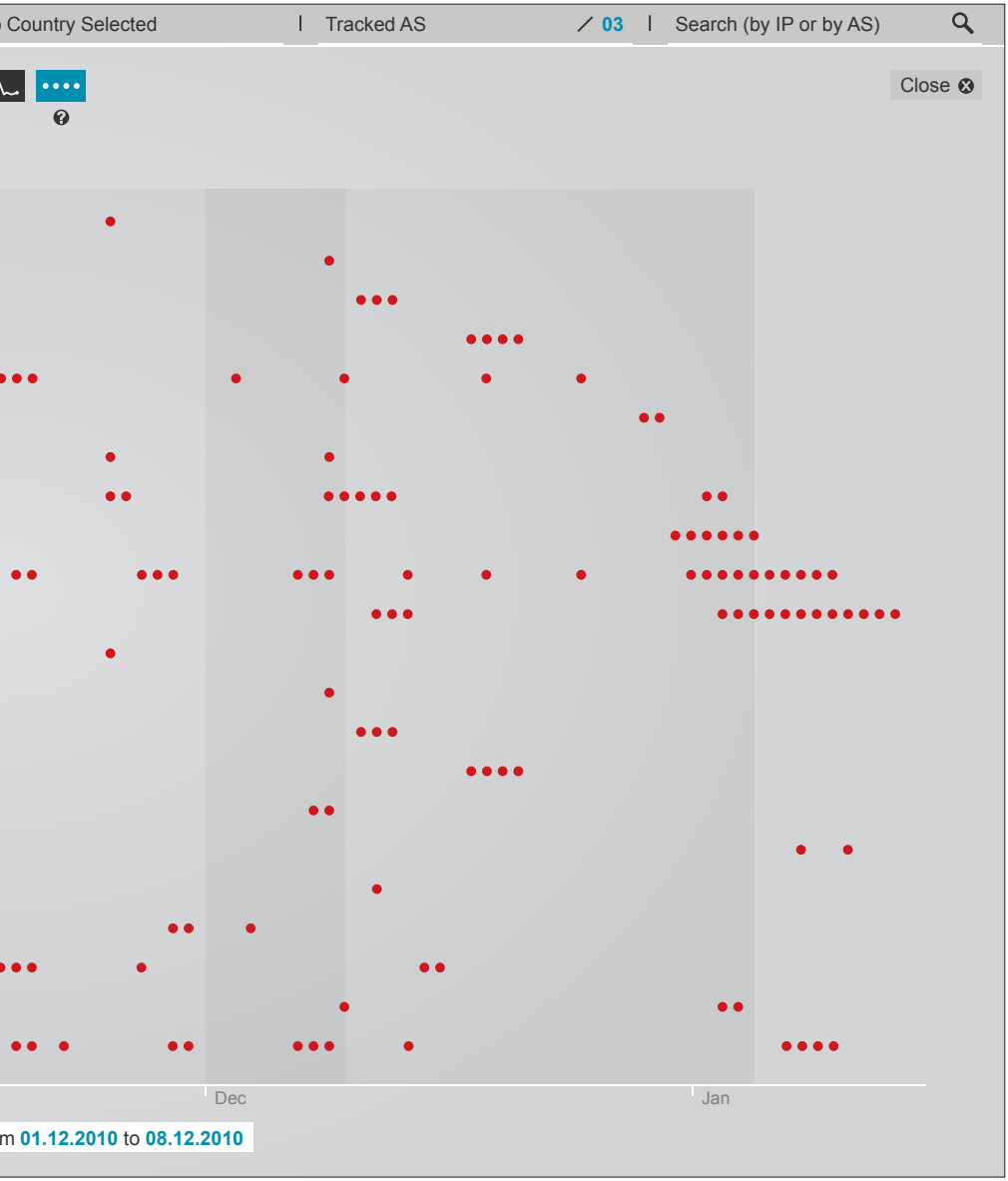
La Service longevity chart consiste di una timeline dove giornalmente viene registrata l'attività di ogni singolo IP, o server, appartenente all'AS. Le righe corrispondono agli IP, mentre le colonne corrispondono ai giorni. Per ogni giorno in cui su di un determinato IP viene registrata attività malevola viene aggiunto un punto rosso sul grafico. In questo modo linee ininterrotte di punti rossi appartenenti ad uno stesso IP evidenziano sul grafico quelli che possono essere definiti come long-living hosts. Nel momento in cui questi ultimi andassero a rappresentare la maggioranza degli IP presenti nell'AS, significherebbe che quella rete non sta prendendo contromisure sufficienti per combattere il fenomeno, o non è interessata a combatterlo. In altre parole quell'AS può essere classificato come tollerante nei confronti delle attività malevole presenti su di esso.

Ogni punto rappresentato sul grafico può appartenere ad una delle quattro diverse tipologie di attività registrate dal sistema (c&c, malware, phishing, spam). Ad un primo livello di lettura non è possibile distinguere tra le varie tipologie di attività, in quanto tutte vengono rappresentate nella medesima maniera. Questo però fornisce un primo grado di informazione, dove l'insieme ci restituisce un'idea dell'andamento generale dell'attività visualizzata. Nel momento in cui si voglia fare una distinzione tra le diverse tipologie, passando in mouse over sui

▶ tav 18 | La service longevity chart

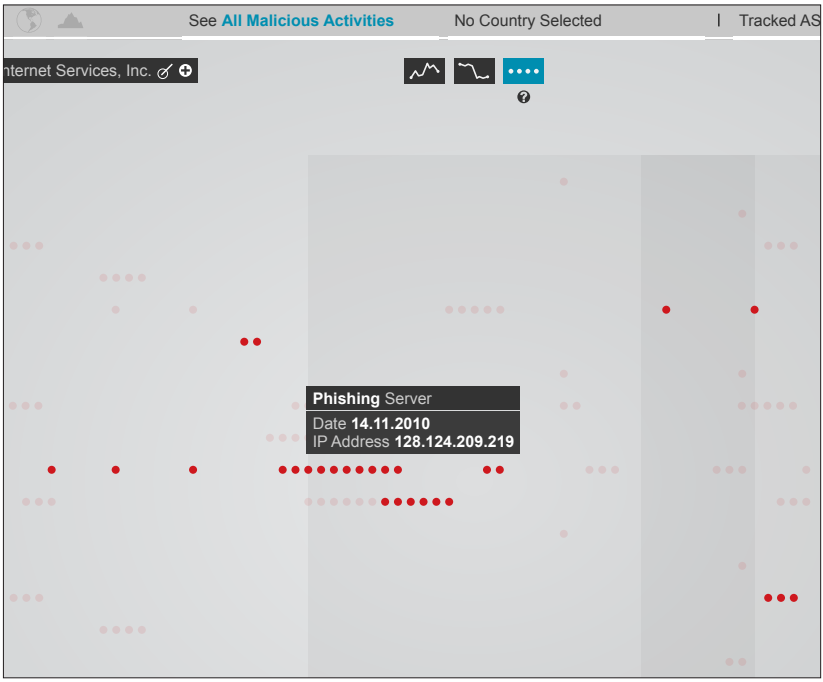


6. BURN, Baring Unknown Rogue Networks


Country Selected
Tracked AS
03
Search (by IP or by AS)
Close
Dec
Jan
m 01.12.2010 to 08.12.2010

singoli punti rossi si andranno ad evidenziare sulla mappa tutti i punti relativi alla stessa tipologia di attività malevola, mandando in opacità gli altri, ed aprendo, come già visto per la history e la trend chart, pop-up informativi. In questo modo attraverso l'interazione l'utente accede ad un secondo livello di lettura dove ha la possibilità di analizzare quali tipologie di attività hanno agito su quali server e quando. Inoltre a questo secondo livello è anche possibile individuare se un server ha avuto uno switch di attività, passando, senza interruzioni da un giorno all'altro, ad esempio da un'attività di spam ad una di phishing.

Sullo sfondo questa volta sono presenti due fasce. La fascia più stretta ci indica qual'è il periodo di attività che abbiamo selezionato attraverso la timeline nella global view, mentre la fascia più ampia indica il periodo di tempo che BURN tiene in considerazione per valutare se l'AS è da considerarsi come tollerante.



▲ **tav 19 | Mouse over service longevity chart**

6.3.3 Service migration screen

La Service migration screen serve a visualizzare le relazioni e gli spostamenti di attività malevola da una rete all'altra, evidenziando le connessioni tra le reti che hanno registrato un improvviso decremento di attività, shutdown, e quelle reti che, nei giorni immediatamente successivi, registrano un improvviso incremento di attività. Per fare questo BURN si avvale di una serie di algoritmi, creati appositamente, che analizzano la quantità e la tipologia delle attività coinvolte in una migrazione, per offrire all'utente una lista di possibili mete degli spostamenti di attività, ordinati secondo un valore di compatibilità.

Non esistendo la certezza matematica relativa ad uno spostamento di attività malevola, la strada che abbiamo seguito nel nostro progetto è stata quella di sfruttare una serie di parametri per creare una lista delle possibili migrazioni di attività, dove le migrazioni con una più alta probabilità di esattezza vengono messe in risalto rispetto alle altre. Anche in questo caso l'analisi manuale da parte dell'uomo si rivela fondamentale per tentare di capire ed indagare a fondo il problema, il nostro sistema offre però un valido strumento a supporto di questa tipologia di ricerca.

Cliccando sul bottone per accedere alla Service migration screen l'utente accederà alla visualizzazione in due fasi:
- in una prima fase, una schermata intermedia gli proporrà, attraverso alcune miniature, una serie di grafici di shutdown, suddivisi tra quelli avvenuti nel periodo che si sta analizzando e altri avvenuti in altri periodi di tempo. Le due categorie di grafici sono raccolte in due sezioni differenti della schermata e sotto ogni miniatura è presente la data in cui lo shutdown è stato registrato. Attraverso questa schermata intermedia l'utente può scegliere quale shutdown andare a visualizzare;
- selezionando lo shutdown l'utente verrà quindi proiettato sulla visualizzazione vera e propria.

▶ tav 20 | La service migration screen: schermata intermedia
▶▶ tav 21 | La service migration screen


BURN
See All Malicious Activities
No
AS21844 THEPLANET-AS - ThePlanet.com Internet Services, Inc.
Shutdown in the selected period
08.12.2010
05.12.2010
Other shutdown occurred on AS21844
22.11.2010
14.11.2010
02.08.2010
11.05.2010


Country Selected
Tracked AS
03
Search (by IP or by AS)
Close
07.09.2010
16.08.2010
03.02.2010
23.12.2009
1/3

6. BURN, Baring Unknown Rogue Networks

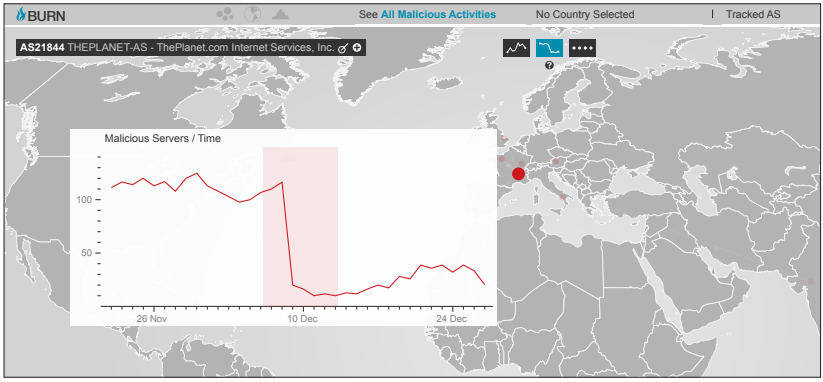

Country Selected
Tracked AS
03
Search (by IP or by AS)
Close
14.11.2010
07.09.2010
16.08.2010

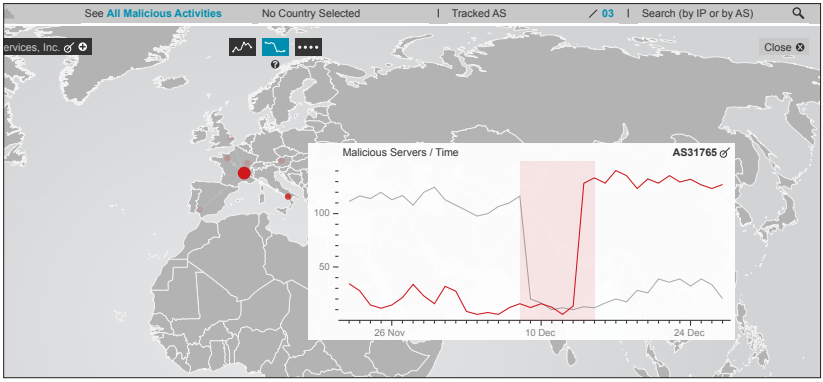
La visualizzazione relativa alla Service migration screen si compone di una mappa geografica, simile a quella già precedentemente descritta, dove l'AS in shutdown e gli AS destinatari di una possibile migrazione vengono posizionati e messi in relazione tra loro. In questo caso tutti gli AS sono rappresentati sotto forma di piccoli cerchi non animati. Dall'AS in shutdown, rappresentato da un cerchio di dimensioni leggermente maggiori rispetto agli altri, partono una serie di linee che lo connettono con le altre reti posizionate sulla mappa. Ogni rete rappresenta una possibile meta di migrazione dell'attività malevola, proveniente dall'AS in shutdown.

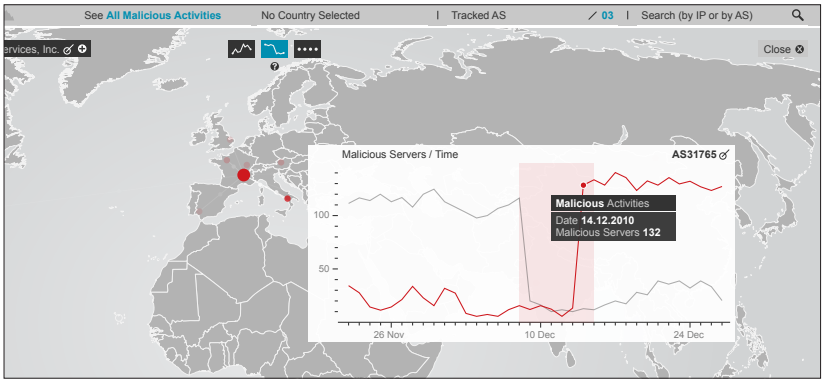
Le reti destinatarie dell'attività, le possibili migrazioni, sono ordinate in base al loro grado di compatibilità, reso visivamente attraverso due fattori:
- più la probabilità che la rete sia effettivamente destinataria dell'attività malevola è alta più il cerchio rappresentante quell'AS sarà di colore intenso. Abbassandosi la probabilità si abbassa anche l'opacità del colore;
- sulle linee che connettono i vari AS viaggiano una serie di impulsi. Anche in questo caso la frequenza degli impulsi è relativa al grado di compatibilità della migrazione, una frequenza di impulsi più alta indica una maggiore probabilità di migrazione.

Passando il mouse sopra i cerchi rappresentanti gli AS si vanno ad aprire dei box in pop-up che permettono la visualizzazione del grafico relativo all'attività indicata. In particolare aprendo il box appartenente all'AS in shutdown si andrà ad aprire un pop-up dove lo shutdown in questione viene rappresentato attraverso un grafico che visualizza il numero di malicious servers registrati su quell'AS nel tempo. Aprendo il box appartenente ad uno degli AS destinatari dell'attività si aprirà invece un pop-up contenente un grafico multiplo dove la linea di andamento dell'incremento di attività relativa all'AS in attivazione viene sovrapposta alla linea di andamento dell'attività relativa all'AS in shutdown. In questo modo viene fornito all'utente un semplice ma efficace mezzo visuale per confrontare tra loro i due eventi e valutare manualmente l'esattezza delle proposte fornite dal sistema. Anche da questi box è possibile aggiungere gli AS alla Autonomous system tracking list.

▶ **tav 22a-b-c | Visualizzazione dei grafici nella service migration screen**


BURN
See All Malicious Activities
No Country Selected
Tracked AS
AS21844 THEPLANET-AS - ThePlanet.com Internet Services, Inc.
Malicious Servers / Time
100
50
26 Nov
10 Dec
24 Dec


See All Malicious Activities
No Country Selected
Tracked AS
03
Search (by IP or by AS)
Close
Malicious Servers / Time
AS31765
100
50
26 Nov
10 Dec
24 Dec

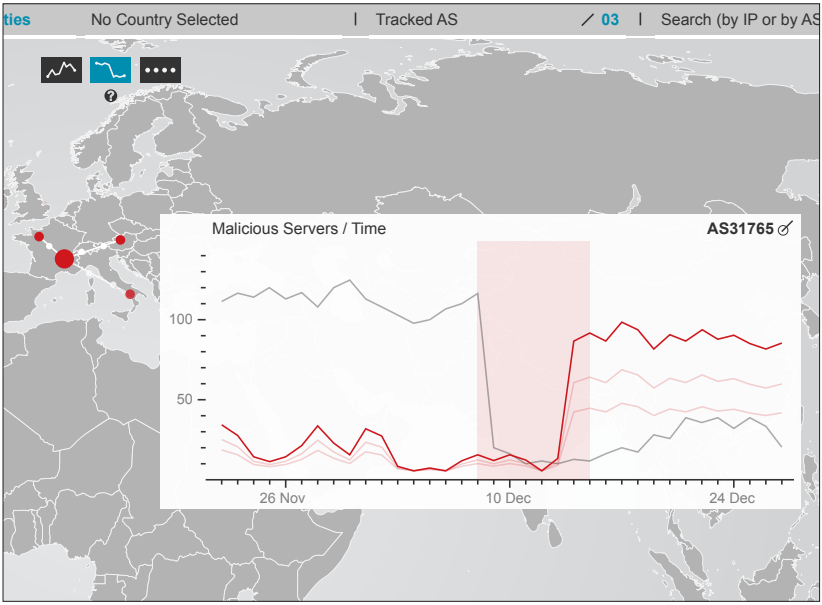

See All Malicious Activities
No Country Selected
Tracked AS
03
Search (by IP or by AS)
Close
Malicious Servers / Time
AS31765
Malicious Activities
Date 14.12.2010
Malicious Servers 132
100
50
26 Nov
10 Dec
24 Dec

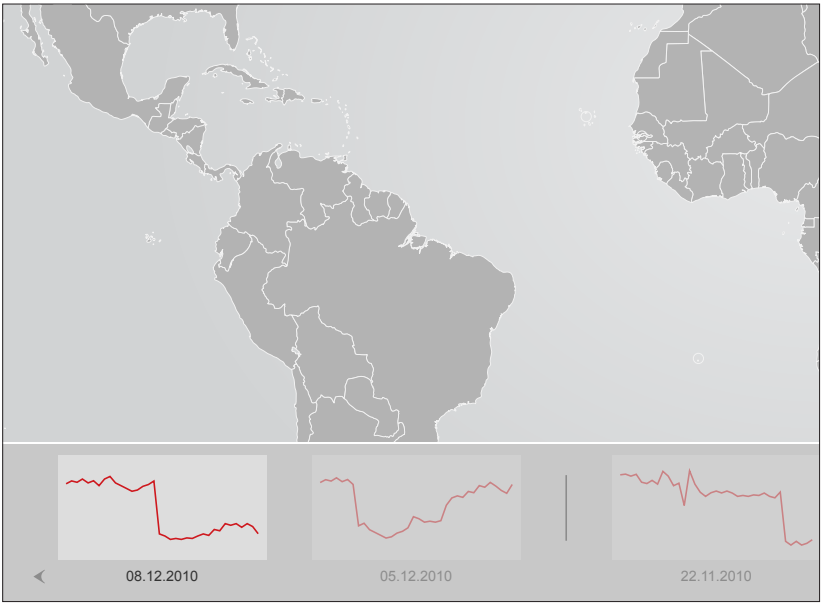
Inoltre vi è la possibilità che una migrazione comporti uno spostamento di attività da una singola rete a molte reti, invece che da una singola rete ad un'altra singola rete. In questo caso passando il mouse sopra il cerchio rappresentante una delle reti in cui l'attività è migrata, nel grafico del box che si apre in pop-up saranno presenti più linee, corrispondenti alle attività dei diversi AS verso cui la migrazione è presumibilmente avvenuta, e sulla mappa rimarranno evidenziati tutti gli AS interessati, permettendo in questo modo una rapida analisi di come la migrazione si sia ripartita tra i vari destinatari.

Nella parte inferiore della visualizzazione un box contenente la lista di tutti gli altri shutdown avvenuti sull'AS, ordinati cronologicamente, permette di passare rapidamente alla visualizzazione delle altre migrazioni. Solitamente visualizzato nella sua forma compressa, passandoci sopra in mouse over si espande per mostrare un'anteprima dei grafici degli shutdown.

▶ **tav 23 | Esempio di grafico di una migrazione multipla**

▶▶ **tav 24 | Lista degli shutdown espansa**


ties
No Country Selected
Tracked AS
03
Search (by IP or by AS
Malicious Servers / Time
AS31765
100
50
26 Nov
10 Dec
24 Dec


08.12.2010
05.12.2010
22.11.2010

6.4 Personalizzare l'interazione coi dati

Attraverso una serie di strumenti e filtri inseriti nell'applicazione, è possibile da parte dell'utente operare alcune importanti personalizzazioni sulle varie visualizzazioni, per personalizzare al bisogno le proprie ricerche.

6.4.1 Timeline e selezione del time range

Di default all'apertura del sistema BURN visualizza i dati relativi all'ultima settimana di attività malevola. La timeline, posta nella parte inferiore di due delle visualizzazioni appartenenti alla global view, bubble chart e geographical map, serve a modificare il periodo di tempo di cui si vogliono visualizzare i dati, e si presenta in due forme: compatta ed estesa.

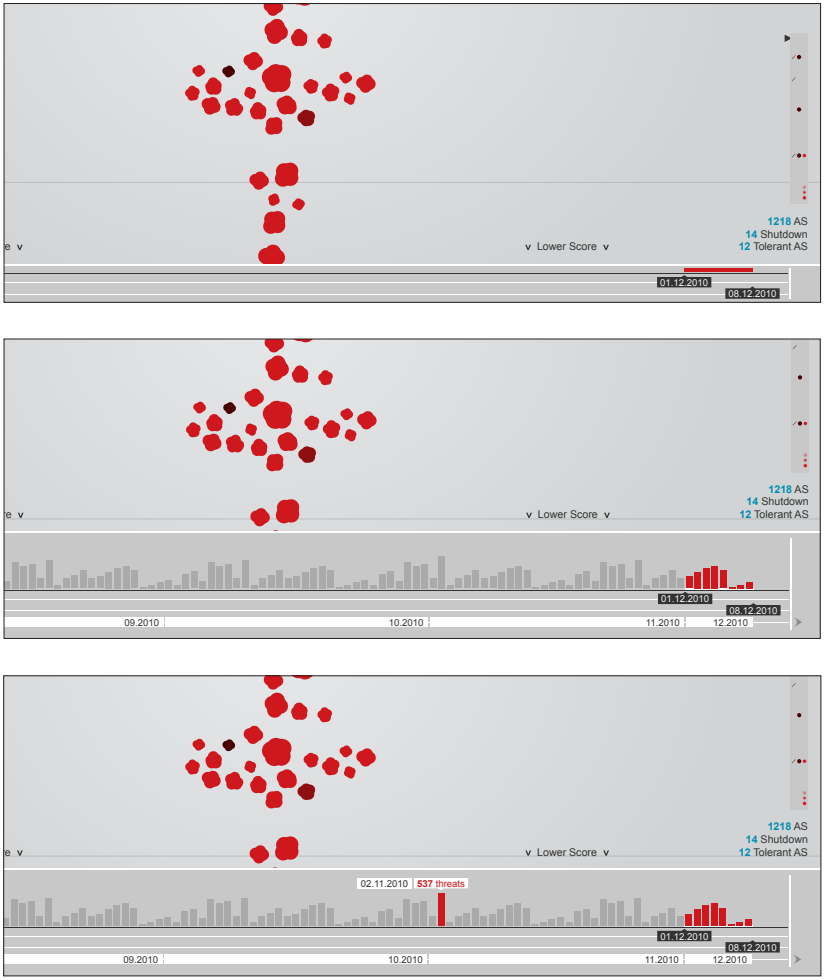
Nella sua forma compatta, realizzata per garantire in ogni momento alle visualizzazioni soprastanti il maggior spazio visivo disponibile, indica, attraverso due etichette, gli estremi del periodo di tempo che si sta visualizzando. Passando in mouse over su un qualsiasi punto della timeline compatta essa si espanderà, per fornire all'utente tutto lo spazio necessario al suo utilizzo. Nella sua forma espansa la timeline è composta da più elementi. Partendo dal basso incontriamo in quest'ordine:

- una riga in cui ci viene fornita una suddivisione in mesi, affiancata a destra e a sinistra da due frecce che consentono lo scorrimento orizzontale di tutta la timeline, per andare a selezionare periodi anche lontani nel tempo (la timeline visualizza in un'unica schermata fino ad un massimo di circa quattro mesi);

- due righe adibite allo scorrimento delle etichette che delimitano il periodo di tempo che si vuole andare a prendere in analisi. La riga superiore contiene l'etichetta relativa al giorno di inizio del periodo, mentre quella inferiore l'etichetta relativa al giorno di fine del periodo. Il periodo di tempo preso in esame può variare da un minimo di un giorno a diversi mesi (tutti i giorni per cui sono presenti dati nel database su cui BURN si appoggia);

▶ **tav 25a-b-c | La timeline: compatta, espansa e mouse over sui singoli giorni**

- un grafico a barre dove ad ogni barra corrisponde un singolo giorno di attività, e la cui altezza è proporzionale al numero totale di server malevoli registrati a livello mondiale. Questo fornisce all'utente un primo colpo d'occhio sui periodi che si stanno andando a selezionare. Passando il mouse sulle singole barre si aprono dei piccoli pop-up contenenti la data a cui la barra fa riferimento ed il numero di server malevoli registrati quel giorno.



6.4.2 Activity filter

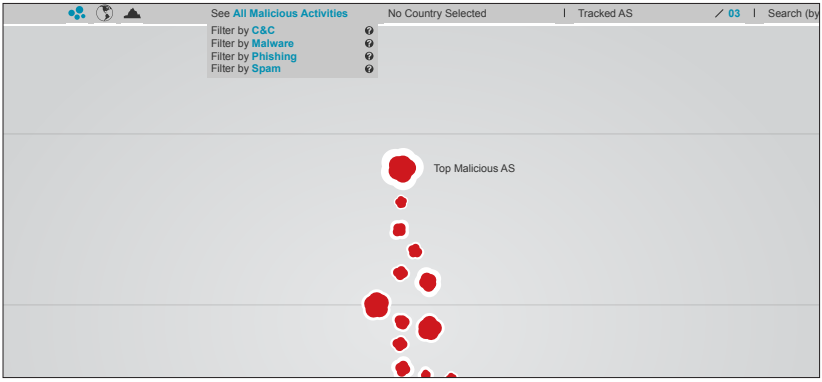
L'activity filter permette di selezionare, attraverso un menù a tendina, la tipologia di attività malevola che si vuole visualizzare: C&C (Botnet), Malware, Phishing, Spam (per una descrizione delle diverse tipologie di attività fare riferimento al capitolo 3, nel paragrafo dove viene spiegato il funzionamento di FIRE). Di default all'apertura del sistema questo menù è posizionato su "See All Malicious Activities", il che significa che in tutte le visualizzazioni presenti si sta visualizzando la situazione generale delle quattro tipologie di attività sommate tra loro. Nel momento in cui l'utente utilizza questo filtro tutte le visualizzazioni, ed anche la timeline, vengono filtrate per l'attività selezionata.

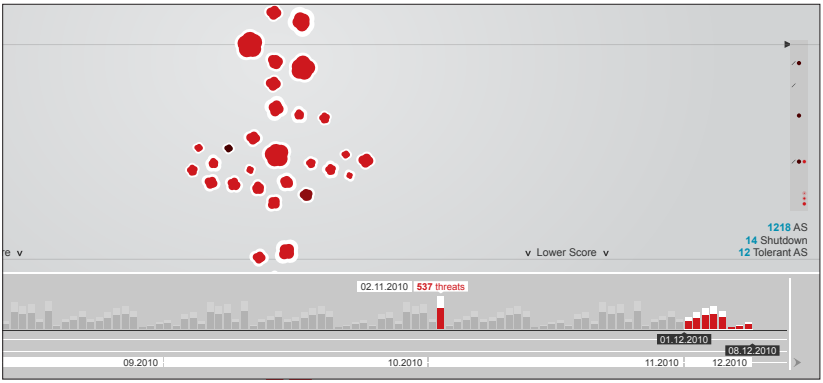
Filtrando le visualizzazioni per una specifica tipologia di attività, oltre a rendere le informazioni relative alla scelta fatta, permette di evidenziare la percentuale che ogni tipologia di attività copre sul totale delle attività registrate. Difatti nella bubble chart, nella mappa geografica e nella timeline viene sempre mantenuto visibile sullo sfondo la forma relativa al totale delle attività malevola, mentre in primo piano si vede la forma relativa alla singola tipologia scelta.

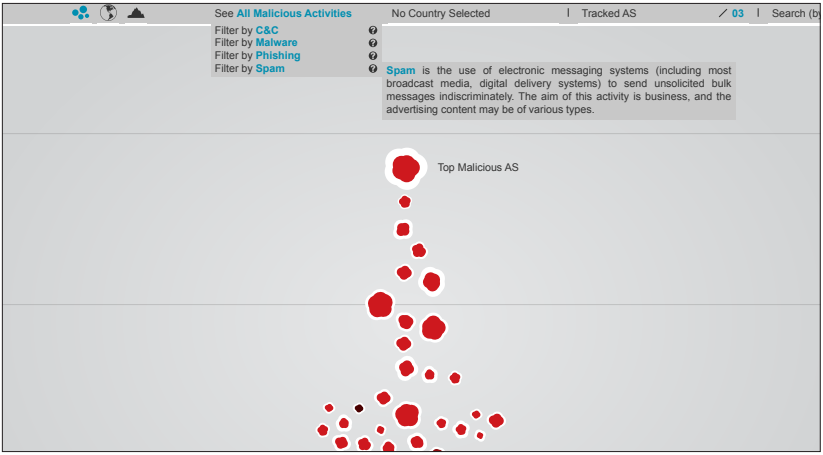
Affianco ad ogni tipologia di attività è inoltre presente un piccolo punto di domanda. Questi bottoni, presenti in molte zone dell'applicazione, permettono, se cliccati, di avere maggiori informazioni, o semplicemente una spiegazione, su ciò che si sta vedendo, e sono pensati soprattutto per quegli utenti meno esperti e meno pratici di alcuni termini strettamente legati al campo della sicurezza informatica, che necessitano di un aiuto in più per comprendere al meglio il fenomeno.

▶ **tav 26a-b | Funzionamento dell'activity filter**

▶▶ **tav 27 | Informazioni aggiuntive**

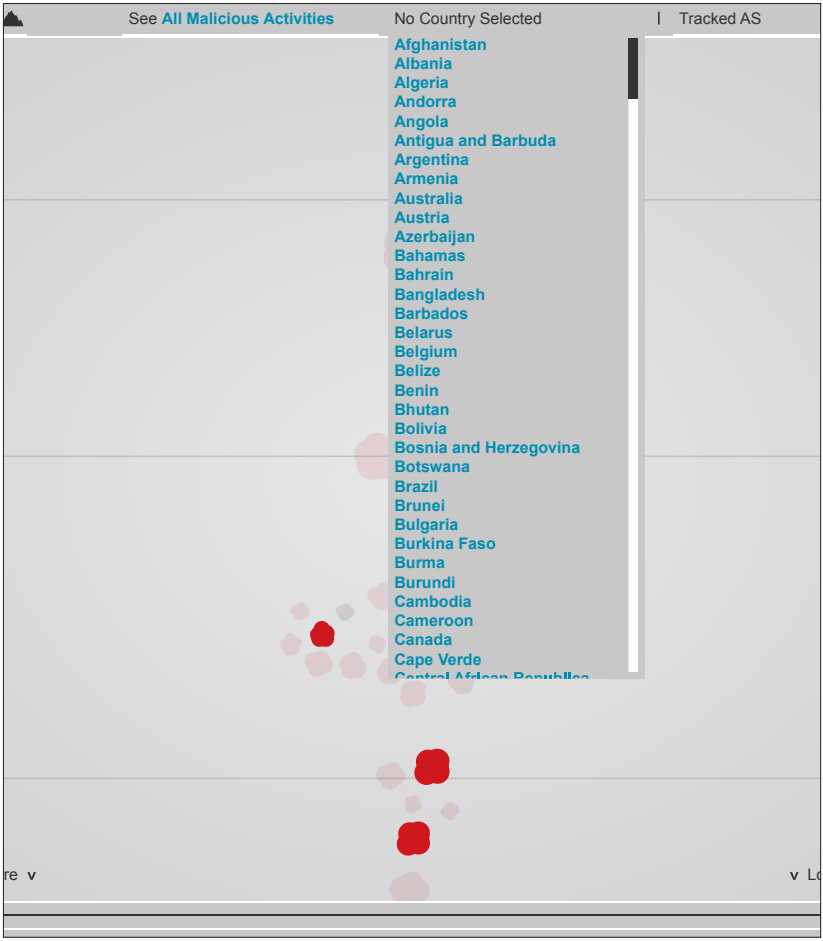

See All Malicious Activities
No Country Selected
Tracked AS
03
Search (by
Filter by C&C
Filter by Malware
Filter by Phishing
Filter by Spam
Top Malicious AS


1218 AS
14 Shutdown
12 Tolerant AS
v Lower Score v
02.11.2010 537 threats
01.12.2010
08.12.2010
09.2010
10.2010
11.2010
12.2010


See All Malicious Activities
No Country Selected
Tracked AS
03
Search (by
Filter by C&C
Filter by Malware
Filter by Phishing
Filter by Spam
Spam is the use of electronic messaging systems (including most broadcast media, digital delivery systems) to send unsolicited bulk messages indiscriminately. The aim of this activity is business, and the advertising content may be of various types.
Top Malicious AS

6.4.3 Country filter

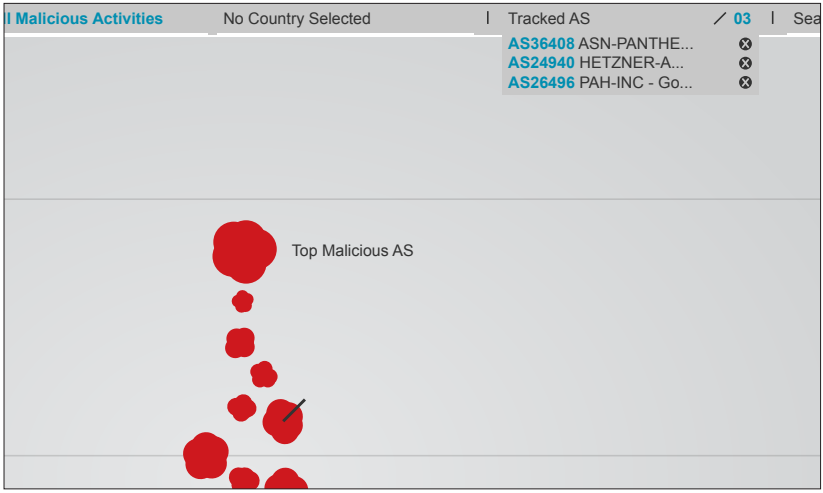
Similmente all'Activity filter, il Country filter permette di selezionare, attraverso un menù a tendina, un singolo stato di cui si vuole monitorare l'attività. Selezionando uno stato attraverso questo filtro tutte le visualizzazioni, compresa la timeline, verranno filtrate mostrando soltanto i dati relativi a quello stato, e la geographical map sarà automaticamente zoommata sullo stato selezionato.



▲ tav 28 | Funzionamento del country filter

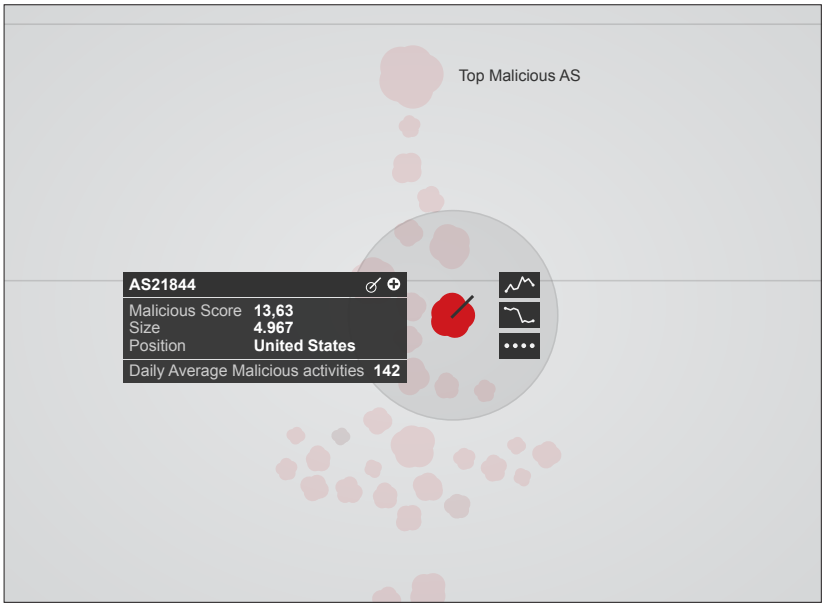
6.4.4 Autonomous system tracking list

Quegli AS che durante l'utilizzo di BURN venissero trovati interessanti dall'utente, possono in ogni momento essere inseriti nella Autonomous system tracking list, attraverso il marcatore presente nei box contestuali che si aprono quando si seleziona un AS, posizionato accanto all'icona per espandere il box. In questo modo gli AS marcati vengono inseriti nel menù a tendina, chiamato "Tracked AS", posizionato tra il Country filter ed il Search, sempre visibile in ogni sezione del sistema. Inoltre, inserendo un AS in questa lista, la bubble relativa alla rete marcata verrà segnata visivamente con una linea diagonale che parte dal centro di essa, in maniera tale da facilitarne il riconoscimento tra le altre, anche nel momento in cui l'utente decidesse di cambiare visualizzazione o alcuni parametri di filtraggio dei dati. Questo strumento risulta molto utile per prendere nota degli AS interessanti in cui l'utente può imbattersi durante le sue ricerche, e crea un importante ponte di collegamento e continuità tra le diverse schermate e visualizzazioni tra cui ci si può muovere utilizzando BURN. Per rimuovere un marcatore da un AS è sufficiente rimuovere quest'ultimo dalla lista presente nel menù a tendina, oppure "spegnere" il marcatore dal box contestuale del singolo AS.



▲ tav 29 | La autonomous system tracking list

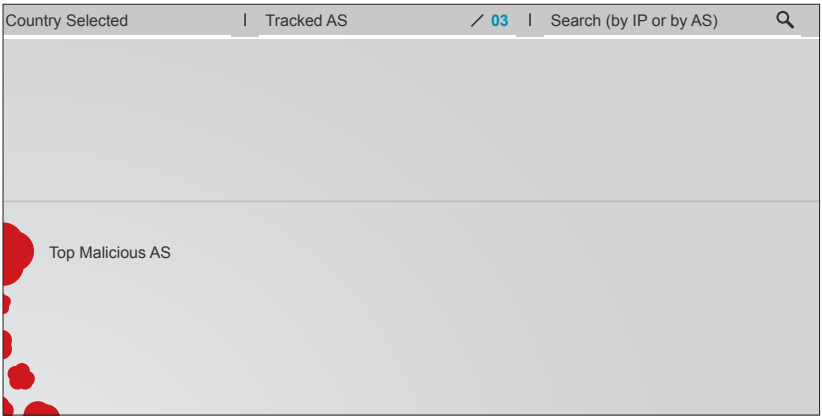
▶ tav 30a-b | Esempio di come vengono marcati gli AS



6. BURN, Baring Unknown Rogue Networks

6.4.5 Search

Nell'angolo in alto a destra dell'applicazione è sempre presente un box per la ricerca. Quest'ultima può avvenire tramite indirizzo IP o numero identificativo di un AS. Inserendo uno di questi due oggetti il sistema selezionerà automaticamente sulla visualizzazione attiva l'AS a cui l'indirizzo IP od il numero identificativo appartiene. In questa maniera è possibile trovare facilmente un oggetto di cui si conosce l'esistenza, verificare a quale AS appartiene un determinato indirizzo IP, e così via.



▲ tav 31 | Il box per la ricerca

6.5 Dettagli sulle tecniche di visualizzazione

In termini di visualizzazione, l'obiettivo principale con cui ci si è voluti misurare durante la progettazione di BURN è stato quello di tentare di superare i limiti dei più comuni modelli di visualizzazione basati sulla codifica delle più tradizionali variabili visive (come colore, forma, dimensione). Si tratta di fatto di modelli di sicura validità, a cui però l'abbinamento di strutture visuali complesse, layout animati interattivi basati su principi fisici e visualizzazioni dinamiche, può essere di importante aiuto nei processi di analisi e pattern recognition (analisi ed identificazione di pattern all'interno di dati grezzi), come nel favorire la comprensione della struttura complessiva del fenomeno e nella generazione di nuove intuizioni e spunti.

6.5.1 L'uso del colore

A differenza di molti sistemi creati ed utilizzati nell'ambito della security visualization, che utilizzano diversi colori per rappresentare diversi "gradi" di sicurezza o insicurezza dei sistemi, gradi di rischio, tipologie di attacchi, e così via, in BURN abbiamo deciso di utilizzare un singolo colore, il rosso, principalmente per tre ragioni:

- per prima cosa, lo spettro dei colori occupato da questa tonalità viene spontaneamente associata ad una sensazione di rischio, di pericolo, cosa che ben si presta per l'argomento qui trattato;
- secondariamente, l'utilizzo di diversi colori (ad esempio per indicare le diverse tipologie di attacco visualizzate) avrebbe potuto portare in errore l'utente, spingendolo, consciamente od inconsciamente, ad associare i diversi colori a diversi gradi di pericolosità od importanza;
- infine, l'utilizzo di un unico colore ci ha dato la possibilità di ridurre in maniera significativa molte delle problematiche inerenti a problemi legati alla visione, come ad esempio il daltonismo o la cecità ai colori.

6.5.2 Le animazioni

In questo progetto abbiamo inserito una serie di animazioni, che agiscono sulle bolle animate presenti nella bubble chart e nella geographical map (quando al maggiore livello di zoom), per i seguenti motivi:
- arricchire le due mappe sopra citate con variabili visive che permettano l'aggiunta di informazioni senza compromettere la leggibilità delle mappe stesse;
- inserire elementi che trasmettano all'utente sensazioni visive strettamente legate a ciò che si sta osservando (ad esempio un elemento che si muove freneticamente può trasmettere una sensazione di elevata attività);
- attirare l'attenzione dell'utente sui fenomeni interessanti che altrimenti rischierebbero di perdersi nella "massa" delle informazioni.

Per l'inserimento delle animazioni è stata condotta una breve ricerca sperimentale, associando alle bubble svariati comportamenti, per studiare quali reazioni e soprattutto quali sensazioni queste suscitassero negli utenti. Questa breve ricerca, incrociata con la necessità di utilizzare variabili che non interferissero tra loro, poiché i comportamenti rappresentati tramite le animazioni non si escludono l'un l'altro ma potrebbero comparire sulla stessa bubble anche tutti contemporaneamente, ha portato alla scelta di utilizzare animazioni che agissero su variabili differenti. Più precisamente sono state utilizzate tre animazioni, che agiscono rispettivamente sulle variabili di forma, dimensione e colore. La prima è sempre attiva, mentre la seconda e la terza agiscono come filtro visivo e si attivano per evidenziare la presenza di un determinato fenomeno.

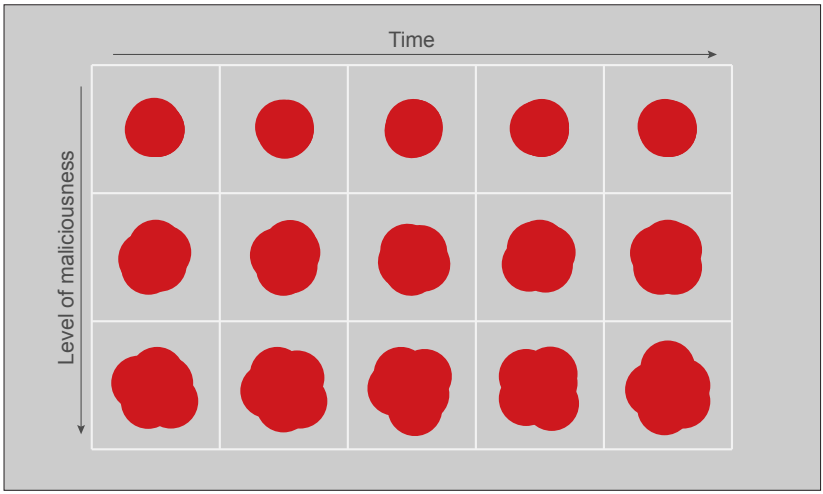
Maliciousness Per monitorare il valore di malicious score di ogni AS rappresentato e, allo stesso tempo, trasmettere all'utente una sensazione di malevolenza legata all'intensità dell'attività, ogni bolla rappresentante un AS è formata da una serie di cerchi sovrapposti che si spostano di continuo nello spazio, sia in verticale che in orizzontale, con un valore di spostamento dal centro dell'AS proporzionale al valore di malicious score assegnato alla rete. In altre parole, più le bubble vibrano più l'AS che rappresentano ha un valore di malicious score alto. Come mostrato nella tavola 32a, questo effetto mira a trasmettere la sensazione che ogni AS sia composto da più elementi, vivi e attivi (i server che compongono la rete). Questa animazione è presente su tutte le bubble presenti nella bubble

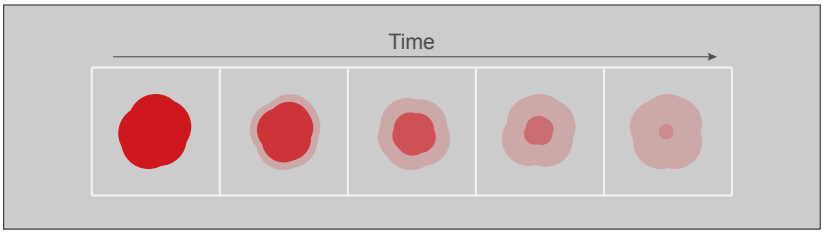
chart e nella geographical map (quando zoommata sul singolo stato), e, se nella visualizzazione di default della bubble chart (dove le bubble sono ordinate per malicious score sull'asse verticale) può risultare come informazione ridondante (ma in ogni caso enfatizzante della situazione), può risultare invece molto utile per individuare quegli AS caratterizzati da una maggiore attività nella geographical map e nella stessa bubble chart quando viene ordinata secondo le altre due variabili disponibili: increments o As size (come precedentemente spiegato nel capitolo relativo alla bubble chart).

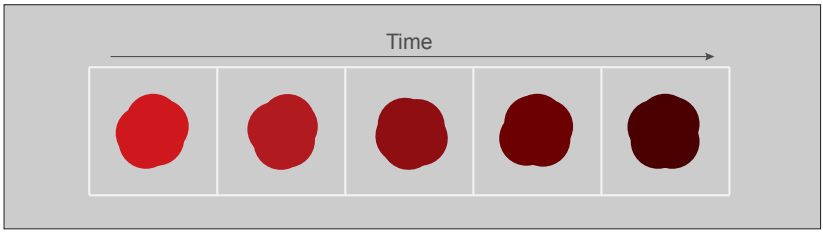
Shutdown Gli AS che, nel periodo di tempo selezionato tramite la timeline, risultano come in shutdown sono visivamente rappresentati da una graduale diminuzione della dimensione e dell'opacità della bubble, fino alla sua scomparsa. In questo caso la diminuzione di attività dovuta ad uno shutdown della rete viene reso visivamente attraverso una diminuzione della variabile che sulle bubble rappresenta il numero di server malevoli presenti nell'AS. L'animazione lavora in loop continuo in maniera tale da facilitare l'identificazione di questi AS tra tutti gli altri rappresentati. La dimensione originale dell'AS viene sempre mantenuta sullo sfondo dell'animazione.

Tolerance In questo caso l'animazione lavora cambiando la luminosità del colore relativo alle bubble rappresentanti quegli AS caratterizzati da un comportamento considerato tollerante nei confronti delle attività malevole, sempre in riferimento al periodo di tempo selezionato tramite la timeline. La luminosità del colore diminuisce gradualmente, scurendo la bubble, e rendendo visivamente la sensazione di una rete marcia, dove l'attività malevola viene tollerata, senza prendere le necessarie contromisure per mitigare la situazione. Anche in questo caso l'animazione lavora in loop continuo.

▶ **tav 32a-b-c | Animazioni: Maliciousness, Shutdown, Tolerance**


Time
Level of maliciousness


Time


Time

6.6 Alcuni scenari di utilizzo

Per esplicitare meglio alcune delle funzionalità permesse da BURN abbiamo sviluppato tre ipotetici scenari di utilizzo.

6.6.1 Trovare, in un periodo di tempo dato, gli AS più malevoli

In questo primo scenario, obiettivo dell'analista è quello di esaminare gli AS che hanno raggiunto un alto punteggio di malicious score, nel periodo di massima attività dell'anno 2010. Per fare questo si andrà a trovare prima l'AS con il maggior punteggio di malicious score sulla somma di tutte le attività, e, successivamente, gli AS suddivisi per le diverse tipologie.

Accedendo a BURN, l'applicazione visualizza l'ultima settimana di attività attraverso la bubble chart. Per andare a trovare picchi di attività malevola all'interno dell'intero anno, l'utente cambia visualizzazione, spostandosi sulla trend chart, che mostra gli andamenti globali dell'attività malevola. In questa visualizzazione è facile individuare un picco di attività tra la fine del mese di maggio 2010 e la prima metà del mese di giugno dello stesso anno. Tornando sulla bubble chart ed agendo sulla timeline viene velocemente selezionato il periodo di tempo interessato, e la bubble chart si aggiorna secondo i nuovi dati forniti, mostrando, come primo risultato a partire dall'alto, l'AS con il più alto livello di malicious score nel periodo selezionato. Cliccando sulla bubble rappresentante l'AS l'analista lo identifica come l'AS21844, ed accede ad alcune informazioni aggiuntive presenti nel box contestuale. Cliccando sull'icona presente nel box contestuale l'AS viene inserito nella lista dei "Tracked AS". Anche il secondo, il terzo, ed il quarto AS vengono aggiunti alla lista. Dopodiché, attraverso i bottoni apparsi cliccando sulla bubble, l'analista accede alla history dell'AS21844. Finita l'analisi di questo AS, rimanendo nella sezione history, attraverso il menù a tendina della lista "Tracked AS", l'analista seleziona gli altri AS che si era segnato, ed accorgendosi così che il terzo AS segnato, l'AS21740, è caratterizzato da un elevato numero di server implicati in attività di C&C. Per avere ulteriore conferma di ciò che ha appena notato, l'analista torna alla bubble chart e, utilizzando l'activity filter, filtra l'attività visualizzata per C&C, e utilizzando il bottone apposito riordina la mappa in base all'attività appena selezionata. Questa azione porta ad un aggiornamento dei dati visualizzati e l'AS21740 guadagna la prima posizione, come

maggiore AS implicato in attività di tipo C&C. Selezionando dall'activity filter anche le altre tipologie di attività presenti, l'analista può velocemente trovare gli AS maggiormente implicati in ogni tipologia di attacco.

6.6.2 Tracciare un attacco

Questo scenario mostra come partendo da un evento isolato, sia possibile andare ad investigare il contesto nell'ambito del quale l'evento è avvenuto. Nello specifico l'attenzione di un esperto di sicurezza informatica viene attratta da una mail sospettata di contenere un URL di phishing. L'indirizzo IP da cui proviene appartiene ad un server connesso ad un AS ritenuto tollerante, che permette a questi server di continuare a lavorare indisturbati per diversi giorni.

Accedendo al sistema l'analista cambia la visualizzazione iniziale passando alla geographical map, per osservare l'andamento generale dell'ultima settimana di attività. Passando in mouse over ottiene alcune informazioni sui vari stati, per poi dirigersi verso il Search box. Qui inserisce l'indirizzo IP appartenente alla mail ricevuta, e una volta avviata la ricerca la visualizzazione automaticamente zoomma sulla Germania aprendo il box contestuale di un AS collocato sul suo territorio. La bubble appartenente a questo AS si muove in maniera frenetica, sintomo di un alto livello di malicious score associato a questa rete, ma l'attenzione dell'analista è attratta da un altro fenomeno: il colore della bubble cambia in un loop continuo, passando dal rosso acceso che caratterizza tutte le bubble presenti ad un rosso scuro molto vicino al nero. In aggiunta a questo un bottone sulla destra della bubble lampeggia leggermente, suggerendo l'apertura della sezione relativa alla Service longevity chart. In questa visualizzazione i long-living host sono caratterizzati da una sequenza di punti rossi adiacenti tra loro, e, come sospettato, l'analista, una volta trovato l'IP in questione, passando in mouse over sui punti rossi che gli appartengono, nota come recentemente il server in questione stia esibendo un'attività di phishing, mentre fino a qualche tempo prima si occupava principalmente di C&C.

6.6.3 Trovare ed evidenziare possibili migrazioni di attività

In questo ultimo breve scenario, un addetto alla sicurezza informatica sta monitorando l'attività di una botnet implicata in un ampio numero di attacchi

di tipo DDoS. L'addetto sa che la botnet è controllata da svariati bot masters localizzati in un AS conosciuto, l'AS36536. Tuttavia questi server sono stati recentemente riportati come inattivi. L'obiettivo dell'addetto è di valutare la correttezza di questi report.

Differentemente dallo scenario precedente, dove la search box veniva usata dalla geographical map, in questo caso l'AS viene cercato a partire dalla bubble chart. Ricercando l'AS attraverso lo strumento di search la mappa scorre fino a posizionare l'AS cercato al centro dello schermo, ed il box contestuale viene aperto. Un'animazione sulla bubble ed il bottone relativo alla Service migration screen suggeriscono di aprire questa sezione. L'addetto, cliccando sull'icona, accede ad una prima schermata intermedia dove gli vengono mostrati vari shutdown avvenuti in diversi periodi di tempo su quella rete. Essendo interessato agli avvenimenti recenti, decide di selezionare il primo shutdown indicato, il più recente avvenuto, aprendo in questo modo la Service migration screen. In questa visualizzazione l'AS è localizzato su di una mappa geografica che ne mostra la posizione e le connessioni di alcune possibili migrazioni di attività. Passando in mouse over sull'AS di partenza l'addetto alla sicurezza ne analizza il grafico dello shutdown, osservando quando il sistema ha iniziato a registrare il drastico calo di attività. Successivamente, spostandosi sugli altri AS rappresentati, i destinatari, ne visiona i grafici relativi, confrontandoli con quello di partenza, per cercare quello che maggiormente potrebbe corrispondere ad una migrazione dell'attività malevola, concentrando la sua attenzione su due AS caratterizzati da un colore più intenso ed una più intensa frequenza degli impulsi, i due AS indicati dal sistema come i più probabili. Passando in mouse over su uno di questi due AS il nostro utente nota come il sistema ogni volta gli apra contemporaneamente i grafici di entrambi, suggerendogli una migrazione verso destinazioni multiple: migrazione confermata dall'alta compatibilità e somiglianza tra l'attività registrata sulla rete di partenza e quella registrata sulle due reti appena trovate, l'AS44050 e l'AS32592.

Conclusioni

Il progetto presentato all'interno di questa tesi ha rappresentato un momento importante di collaborazione e confronto tra due discipline. Durante tutte le fasi del lavoro, da quelle iniziali di analisi e ricerca, fino allo sviluppo finale, ci si è continuamente preoccupati di tenere sotto controllo necessità e obiettivi derivanti da entrambe le realtà coinvolte, cercando di combinare al meglio le esperienze fatte in questi due campi. Grazie a questo continuo scambio di conoscenze ed informazioni, e lavorando a stretto contatto con due realtà dedite alla ricerca universitaria, il VPlab ed il DensityDesign, abbiamo potuto costruire e affinare il nostro progetto passo per passo, affidandoci ai continui feedback diretti derivanti proprio da questa collaborazione. Inoltre, a coronamento di questa esperienza, abbiamo avuto anche la possibilità di provare ad esporre e sottoporre il nostro progetto in due importanti occasioni internazionali, che ci hanno dato la possibilità di confrontarci con esperti direttamente interessati al settore, e con una comprovata esperienza del campo.

La prima opportunità che si è presentata è stata quella di partecipare a VizSec2011, l'ottavo simposio internazionale sulla "visualizzazione per la sicurezza informatica", che si terrà alla Carnegie Mellon University, Pittsburgh, PA, USA, il 20 luglio 2011. Per la partecipazione a questo evento è stato preparato un paper [1], sotto-

[1] F. Roveta, L. Di Mario, F. Maggi, G. Caviglia, S. Zanero and P. Ciuccarelli. BURN: Baring Unknown Rogue Networks. In Proc. of the 8th Intl. Symposium on Visualization for Cyber Security, VizSec '11 (to appear).

posto ed in seguito accettato dalla commissione e pubblicato agli atti di una conferenza internazionale, che oltre all'opportunità di confrontarsi con la stesura di una pubblicazione scientifica, ci ha dato anche la possibilità di avere dei feedback diretti da reviewer esperti del settore. Inoltre il giorno della presentazione rappresenterà un ulteriore importante momento di feedback sul lavoro svolto, permettendoci di avere un confronto diretto con un pubblico di ricercatori, aziende ed esperti del settore, specificatamente interessati all'utilizzo delle tecniche di visualizzazione applicate al campo della sicurezza informatica.

La seconda opportunità che ci si è presentata è stata la possibilità di inserire il nostro lavoro all'interno del progetto di ricerca europeo WOMBAT. Questo progetto, arrivato nella sua fase finale nel mese di giugno del 2011, ha organizzato come incontro di chiusura la presentazione di tutti i lavori di ricerca svolti nel corso dei tre anni della sua durata. Il nostro progetto di tesi, del quale è stata preparata una breve demo dimostrativa, è quindi stato inserito tra i lavori presentati in questo evento, ed anche in questo frangente ci ha dato la possibilità di confrontarci con un feedback diretto da parte di ricercatori ed esperti del settore direttamente interessati alle problematiche da noi affrontate. Inoltre a seguito di questa esperienza è stato anche dimostrato da alcuni dei presenti un forte interesse nei confronti sia del lavoro svolto, sia della metodologia con la quale quest'ultimo è stato affrontato e realizzato, interesse che potrebbe sfociare in alcune interessanti collaborazioni future.

Il lavoro svolto in questo periodo ci ha dato quindi la possibilità di entrare in contatto diretto da una parte con il mondo della ricerca universitaria, e dall'altra con gli ipotetici utilizzatori finali del nostro progetto. Questo ha fatto sì che vi fosse una crescita costante e continua guidata da entrambi questi fattori, la cui compresenza ha contribuito alla creazione di un prodotto il più possibile completo e attento a tutte le necessità, i bisogni e le attese degli attori in campo.

Riferimenti Bibliografici

Sicurezza informatica

Inside Cyber Warfare: Mapping the Cyber Underworld.
Jeffrey Carr, O'Reilly - December 2009.

Fire: Finding rogue networks.
B. Stone-Gross, C. Kruegel, K. Almeroth, A. Moser, and E. Kirda. In Proc. of the 2009 Annual Comp. Security Applications Conference, ACSAC '09, pages 231–240, Washington, DC, USA, 2009. IEEE Comp. Society.

An incident analysis system nicter and its analysis engines based on data mining techniques.
D. Inoue, K. Yoshioka, M. Eto, M. Yamagata, E. Nishino, J. Takeuchi, K. Ohkouchi, and K. Nakao. In Proc. of the 15th intl. conference on Advances in neuro-information processing - Volume Part I, ICONIP'08, pages 579–586, Berlin, Heidelberg, 2009. Springer-Verlag.

Security visualization

Applied Security Visualization.
Raffael Marty, Addison-Wesley - August 2008.

Security Data Visualization: Graphical Techniques for Network Analysis.
Conti Greg, No starch press inc. - September 2007.

Ember: a global perspective on extreme malicious behavior.
T. Yu, R. Lippmann, J. Riordan, and S. Boyer. In Proc. of the Seventh Intl. Symposium on Visualization for Cyber Security, VizSec '10, pages 1–12, New York, NY, USA, 2010. ACM.

Interactive detection of network anomalies via coordinated multiple views.
L. Harrison, X. Hu, X. Ying, A. Lu, W. Wang, and X. Wu. In Proc. of the Seventh Intl. Symposium on Visualization for Cyber Security, VizSec '10, pages 91–101, New York, NY, USA, 2010. ACM.

Interactively combining 2D and 3D visualization for network traffic monitoring.
Erwan L Malécot, Masayoshi Kohara, Yoshiaki Hori, and Kouichi Sakurai. In: VizSEC: Proceedings of the 3rd international workshop on Visualization for computer security. ACM, New York, NY, USA,, pages 123–127. 2006.

Visual analysis of code security.
John R. Goodall, Hassan Radwan, Lenny Halseth. In: VizSec '10 Proceedings of the Seventh International Symposium on Visualization for Cyber Security, 2010. ACM.

Visualizing graph dynamics and similarity for enterprise network security and management.
Qi Liao, Aaron Striegel, Nitesh Chawla. In: VizSec '10 Proceedings of the Seventh International Symposium on Visualization for Cyber Security, 2010. ACM.

BURN: Baring Unknown Rogue Networks.
F. Roveta, L. Di Mario, F. Maggi, G. Caviglia, S. Zanero and P. Ciuccarelli. In Proc. of the 8th Intl. Symposium on Visualization for Cyber Security, VizSec '11 (to appear).

Information visualization

The Visual Display of Quantitative Information – Second Edition.
Edward R. Tufte, Graphics Press – August 2009.

Envisioning Information.
Edward R. Tufte, Graphics Press – November 1990.

Visual Explanations: Images and Quantities, Evidence and Narrative.
Edward R. Tufte, Graphics Press – December 1997.

Sémiologie graphique: Les diagrammes - Les réseaux - Les cartes.
Jacques Bertin, Editions de l'Ecole des Hautes Etudes en Sciences - January 1999.

Elements of cartography – Sixth Edition.
Arthur H. Robinson, Joel L. Morrison, Phillip C. Muehrcke, A. Jon Kimerling, Stephen C. Guptill, John Wiley & Sons, Inc. - 1995.

Information Dashboard Design: The effective visual communication of data.
Stephen Few, O'Reilly – January 2006.

Information Graphics: A comprehensive Illustrated Reference. A visual tool for Analyzing, Managing, and Communicating.
Robert L. Harris, Oxford University Press - 1999.

The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations.
Ben Shneiderman, Visual Languages, IEEE Symposium on - September 1996.

From Data to Knowledge. Visualizations as transformation processes within the Data-Information-Knowledge continuum.
Luca Masud, Francesca Valsecchi, Paolo Ciuccarelli, Donato Ricci, Giorgio Caviglia. Information Visualisation (IV), 2010 14th International Conference. 445 - 449. 26-29 July 2010, London, United Kingdom.

The Information Mural: a technique for displaying and navigating large information spaces.
Jerding, D.F., Stasko, J.T.. In: Visualization and Computer Graphics, IEEE Transactions on. Jul-Sep 1998.

Oltre la metafora del cruscotto. Un modello per l'esplorazione visuale dei dati basato sui flussi.
Michele Mauri, Tesi di laurea magistrale. Politecnico di Milano - 2008/2009.

Ringraziamenti

Desidero ringraziare il professore Paolo Ciuccarelli e tutto il DensityDesign per il tempo dedicatomi, per i consigli e per le conoscenze condivise durante tutto il periodo di sviluppo di questa tesi. Il professore Stefano Zanero ed il VPlab per avermi "adottato" durante il periodo di sviluppo del progetto, e Federico Maggi per il supporto fondamentale che ci ha dato durante tutte le fasi di avanzamento del nostro lavoro. Un grazie a tutti quanti sopra citati anche per aver creduto nelle potenzialità del nostro progetto ed averci dato la possibilità di partecipare a VizSec2011 e di inserire il nostro lavoro all'interno del progetto WOMBAT. Un grazie anche ad Alberto Volpatto ed Alessandro Frossi per averci aiutato con lo sviluppo della demo.

Desidero inoltre ringraziare tutte le persone che nel corso di questi anni hanno vissuto con me questo percorso di studi universitario, compagni ed amici con i quali spero di poter condividere ancora molto, e tutti i docenti che hanno preso parte alla mia crescita professionale.

Infine un grazie alla mia famiglia per aver sempre avuto fiducia in me ed avermi lasciato libero nelle mie scelte, a tutti coloro che mi hanno supportato e sopportato nei momenti di stress, e a Luca che ha condiviso con me l'esperienza di questa tesi.

Francesco Roveta, luglio 2011