**2017 9th International Conference on Cyber Conflict**
Defending the Core
H. Rõigas, R. Jakschis, L. Lindström, T. Minárik (Eds.)


# Core Illumination: Traffic Analysis in Cyberspace

**Kenneth Geers**
Comodo Group
Toronto, Canada

**Abstract:** The information security discipline devotes immense resources to developing and protecting a core set of protocols that encode and encrypt Internet communications. However, since the dawn of human conflict, simple traffic analysis (TA) has been used to circumvent innumerable security schemes. TA leverages metadata and hard-to-conceal network flow data related to the source, destination, size, frequency, and direction of information, from which eavesdroppers can often deduce a comprehensive intelligence analysis. TA is effective in both the hard and soft sciences, and provides an edge in economic, political, intelligence and military affairs.

Today, modern information technology, including the ubiquity of computers, and the interconnected nature of cyberspace, has made TA a global and universally accessible discipline. Further, due to privacy issues, it is also a global concern. Digital metadata, affordable computer storage, and automated information processing now record and analyse nearly all human activities, and the scrutiny is growing more acute by the day. Corporate, law enforcement, and intelligence agencies have access to strategic datasets from which they can drill down to the tactical level at any moment. This paper discusses the nature of TA, how it has evolved in the Internet era, and demonstrates the power of high-level analysis based on a large cybersecurity dataset.

**Keywords:** *traffic analysis, malware, cyber operations, geopolitics*

## 1. INTRODUCTION: TRAFFIC ANALYSIS

Core Internet security protocols, including encryption, protect users from a wide range of threats. However, attackers can use traffic analysis (TA) to defeat almost any level of security precaution, as long as they have visibility and a capability to collect and analyse data. In fact, TA is a necessary precursor to cryptanalysis, and it is where strategic signals intelligence (SIGINT) almost always begins.

Digital TA relies on the examination of network flow and metadata, can be effective even against unbreakable encryption, and is often sufficient to act as a basis for both tactical and strategic intelligence insight. Basic information begins with source and destination addresses, message type, count, timing, frequency, length and other 'externals.' With only these data points, it is possible to deduce a surprising amount of intelligence regarding the communicants, including their identity, location, movement, behaviour, capabilities, intentions and morale.

From a military perspective, TA can determine chain of command, order of battle (OB), security level, indications and warning (I&W) and more. With such intelligence in hand, it may then be possible to jam, censor, or deceive an adversary.

TA even has some advantages over having complete access to the adversary's unencrypted, plaintext messages, including:

- speed – TA can be automated;
- cost – content analysis and language translation capabilities are expensive disciplines for which there is never enough expertise or manpower;[1] and
- surprise – TA yields many discoveries, some of them unexpected.

TA and counter-TA have been used throughout political, military, and economic history. Consider three famous examples from World War II: prior to its surprise attack on Pearl Harbour, the Japanese navy broadcast false in-port communications to fool American eavesdroppers; in Operation Quicksilver, the Allies played a similar game against German intelligence to divert Hitler's attention from Normandy; and in helping to crack the Enigma machine, Alan Turing discovered weaknesses in Nazi communications, first by TA, then by cryptanalysis.

SIGINT and Electronic Warfare (EW) have always been key elements of military planning and operations, and TA has always been a precursor and a critical piece of both SIGINT and EW. For example, in a military setting, the extreme secrecy surrounding submarine deployments means that boat captains must balance the benefit of connecting to the chain of command with the risk of being located by adversary vessels using direction finding. Thus, submarines must follow the strictest communications standards and procedures.

Traffic analysis can be performed on anything, from pizza deliveries at the Pentagon to noting the tail numbers of suspicious airplanes. Law enforcement and counterintelligence routinely tally the electricity use and bill payment methods of suspected criminals and spies. In contrast to military and intelligence agencies, civilian enterprises and individual citizens can be particularly vulnerable to TA, as they may not have adequate (or any) operations security (OPSEC) training or experience.

Any smart TA researcher with a large digital dataset has enormous possibilities, including for tracking advanced persistent threat (APT) actors, and even predicting certain future events. There is a well-understood cyber 'kill chain': phishing, for example, is both a principal component of, and a precursor to, most significant cyber attacks. Likewise, the prepositioning of hacker tools on industrial control systems may be considered a latent national security threat.

1 Once a target has been selected at the strategic level for closer scrutiny at the tactical level, more expensive resources can be used: anything from computer hacking to human surveillance and physical destruction.

TA can place these cyber incidents in a wider context, and make them understandable even to non-technical decision-makers.

TA enhances traditional malware analysis. Take Stuxnet or the Democratic National Committee hack; no one researcher, company, or even nation provided us with our understanding of these attacks. Strategic analysis and insight stemmed from the work of hundreds of researchers in different countries – most of whom did not know each other personally – performing not only tactical malware analysis but strategic TA as well. TA can help chip away at the attribution problem, or the challenge of solving the problem of the 'last hop', by collating data points from many different types of sensors that exist in disparate legal jurisdictions, and can be used to discover operations by even the world's most secretive three-letter agencies.

## 2. LITERATURE REVIEW

One of the author's goals in this paper is to bring the significance of TA to a wider audience, as this topic has historically been confined to relatively small circles of specialised analysts. Another goal is to demonstrate just how quickly large digital datasets which encompass communications from around the world can be leveraged for both tactical and strategic insight.

This section references several dozen papers that discuss TA in a wide variety of settings. For example, TA has a strong history in economics, such as for the relief of road congestion [1] [2]. Digital TA dates from at least the early 1990s [3], with one company offering commercial counter-TA solutions as early as 2000 [4].

Digital TA has been performed at every scale, from the size of a computer chip, where the evaluation is mostly physics [5], to botnet research that encompasses software behaviour characterisation, honeypots, virtual machines, counterintelligence [6] and natural language analysis in chat rooms as a form of Turing Test [7]. TA covers both the temporal [8] and spatial examination of data [9].

Some authors have written academic TA overviews that blend historical viewpoints and modern information technology [10] to demonstrate the evolution of the concept. TA is a huge topic, with attacks ranging from time correlation to statistical disclosure and *a priori* knowledge [11]. But the basic idea of TA involves data collection, organising information into network flows [12] and putting it all together into an intelligence framework [13].

Counter-TA is also a mature discipline. Research has focused on how best to protect the location and operational logic of base stations [14, 15, 16], how to disguise network protocols [17], how to falsify traffic, how to move secretly between communications channels [18], how to hide communications within public key cryptography [19], how to pad traffic, how to reroute messages and how to simulate network entropy [20]. Alas, many of these solutions will depend on the quality of the technical personnel and the size of the IT budget.

There are many reasons why counter-TA fails. Effective TA tools can be freely downloaded from the web [21], and highly intrusive TA can be performed remotely. Encrypted tunnelling of HTTP traffic, for example, is vulnerable to TA attacks on time and bandwidth that can identify who you are and where you are going on the Internet [22]. The myriad ways in which TA is possible means that cyberspace is, ultimately, tough terrain for both privacy and human rights [23].

Digital TA begins at the connection level[2] [24], and covers every network protocol, from ADSL [25], to SNMP [26], IPTV [27], peer-to-peer (P2P) [28], DNS, and HTTP [29]. All Internet Service Providers are well-positioned to perform TA against the gamut [30], and as humans begin to live in virtual worlds, TA research has followed them, from Second Life [31, 32] to World of Warcraft [33, 34], and more [35, 36]. For white hat TA, part of the goal is to discover whether another online character is a real human or a game bot [37].

What all this means is that so-called 'anonymous' communications are not as secure as one might think, and when users flock to promising new systems, attackers will turn their sights in that direction [38]. Tor, for example, is a low-latency framework that is considered secure enough for normal web browsing, but likely insecure against TA from a strategic adversary such as a nation-state. These same TA strategies and tactics also work against covert channels [39].

Finally, while it is true that the volume of modern digital communications would seem overwhelming to any analyst, a variety of hardware [40] and software tools [41] have been created for big data analysis, including both licensed and open source network visualisation tools [42, 43].

## 3. TRAFFIC ANALYSIS 2.0

Modern information technology (IT), including the convergence of most communication streams over the same digital networks, has transformed TA. Due to the ubiquity of computers, and the interconnected nature of networks, cyberspace is now a reflection of all human affairs. Everything from politics to romance, business to crime, and espionage to military invasions, can be seen by anyone who has network access and the knowledge to translate Internet protocols into human language.

There are myriad forms of computer hardware and software today, but a basic requirement for interoperability is that most of them must use the same network 'stack', which in turn makes them vulnerable to capture and analysis by eavesdroppers, even when message content is encrypted. For the most part, Internet routing is transparent, and despite the astonishing number of communication devices on the Internet, there are many hardware and software tools that can intercept, store, process and analyse captured data.

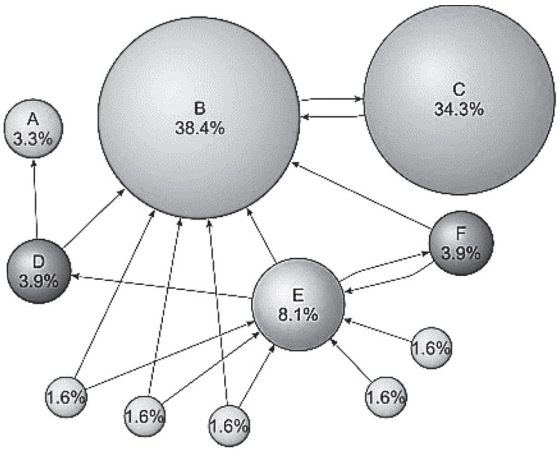
TA is effective at small and large scale. At the micro level, one can identify specific devices[3] and software configurations; it is also possible to map networks by pinging and probing

2 E.g., source IP address, destination IP address, source port number and destination port number.

3 Eavesdroppers can even remotely analyse the 'drift' of a digital clock. Web servers often perform timing measurements in order to make inferences about site visitors.

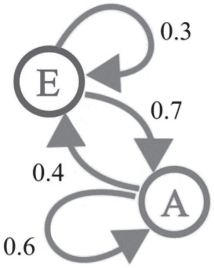
unfamiliar network space.[4] At the macro level, strategic datasets and algorithms are used on a daily basis. For example, the Google PageRank algorithm (see Figure 1) can evaluate the relative significance of web pages across the Internet, by counting the number and quality of hyperlinks pointing to a given site.

**FIGURE 1.** MATHEMATICAL PAGERANK MODEL (SOURCE: WIKIPEDIA)




Markov chain models (see Figure 2) are used to predict where Internet users will go next, based upon where the user currently is, and the use of probability theory.

**FIGURE 2.** MARKOV CHAIN MODEL (SOURCE: WIKIPEDIA)




These and similar TA algorithms are regularly exploited to identify users and predict what they will buy and for whom they will vote.[5]

Core Internet security protocols are also vulnerable to TA. Secure Shell (SSH), which encrypts all communication streams over unsecured networks, is vulnerable to attacks that simply count the number and timing of network packets.[6] Transport Layer Security (TLS) and its predecessor

4 Nmap (Network Mapper) can identify operating systems, open ports, running services and more.
5 Social network analysis was famously used to locate Saddam Hussein via his tribal and family links.
6 Keyboards have fixed layouts that sometimes allow attackers to guess passwords based on the time it takes for human fingers to move between individual keys.

Secure Sockets Layer (SSL) do not obscure communications in a way that conceals their message sizes; this allows TA to discover which webpages a user has accessed.

## *A. Metadata*

All digital communications generate activity records in the form of log files,[7] which can, to a large degree, indicate what takes place in traditional geopolitical space.[8]

Therefore, TA can illuminate physical, terrestrial activity by examining individual log files. However, it is far more effective to concatenate log files from numerous sources, which is akin to having multiple witnesses testify in a court trial. In this way, even the most secure networks are vulnerable to TA. For example, it often happens that a target is invisible on one layer in the network stack, but not on another.

Despite the enormous volume of metadata, TA specialists can automate much of their work, including by using advanced mathematical models that discover previously unknown correlations and anomalies between any two objects in a large dataset. Analysts may seek any number of interesting network patterns, but they often include political, military and economic intelligence.

Again, TA does not demand the availability of any messages in their plaintext (unencrypted) forms. For automation purposes, human conversations are in fact notoriously difficult for computers to understand.[9] Message content is even tricky for human analysts to follow, since many words are culture- and context-specific, and associated with events with which only the communicants are familiar.[10]

In the event that an adversary has employed so many special security protocols that they are nearly invisible, it is also possible that such extreme measures will themselves be discovered as anomalous, and only serve to pique the interest of third parties and raise the level of TA to which any such network is subjected.

## *B. Geolocation*

The implications of digital TA are serious: if researchers or attackers can identify you, they also might be able to find you in the real world. Historically, TA employed radio frequency (RF) direction finding, with a line of bearing to a transmitter. However, today there are newer technologies such as the Global Positioning System (GPS) and Time Difference of Arrival (TDOA), which triangulate network users via satellite and cell phone towers.

Many browser-based tools can plot digital communications on a real-world map.[11] The most common method is to look up an Internet Protocol (IP) address in a Whois database, where

7 Eavesdroppers may find interesting logs anywhere, from browser caches to web server and router logs.

8 It is beyond the scope of this paper, but computing devices also emit physical signatures that can be measured and captured with the right equipment.

9 The difficulty of passing the 'Turing Test,' for example, highlights how difficult it is for computers to appreciate human language in context.

10 This is why, for example, that any type of censorship is difficult to perform accurately, and almost perforce leads to over-censorship.

11 Standards include ISO 3166, ISO/IEC 19762-5:2008, FIPS, INSEE, Geonames, IATA, ICAO, American National Standards Institute (ANSI) Codes, WOEID (Where on Earth IDentifier), NAC Locator, geotagging, location-based services, mobile phone tracking, W3C Geolocation API, geolocation video and more.

anyone can see the IP's registrant, physical address, associated domain names, business name and more.[12] Sophisticated geolocation can be performed on MAC addresses, RFID, embedded code, Wi-Fi positioning systems, device GPS coordinates, archival tags, microchip implants, data storage tags, and more. Using social media, it is also possible to geolocate images, videos, and comments, most of which are self-disclosed. Finally, online search engines and mapping software such as Google Earth can be used to refine and display geo-coordinates.

A recent trend has been the collaborative efforts of cybersecurity researchers and firms worldwide to investigate attacks using a wide range of the free tools described above. For example, Bellingcat and Vice News have tracked Russian military forces in Ukraine; and international, crowdsourced efforts have reverse-engineered major cybersecurity incidents such as Stuxnet and the 2016 attack on the US Democratic National Committee.

## *C. Attribution*

In cyber defence analysis, the biggest challenge is typically attribution, or determining the true source of an attack. This is also known as the problem of the 'last hop', and refers to the known IP address with which a hacker interacts with a victim computer, for reconnaissance or exploitation purposes. As a rule, it is a compromised computer that lies within a broader attack infrastructure, and which the attacker uses as a temporary stepping stone across the Internet.

Strategic TA is one of the primary ways in which cyber defenders can correlate attack data, by connecting disparate communication streams from logs collected at different points on the Internet. This may be done passively by analysing the log files, or actively, by, for example, creating honeypots that place homing beacons on stolen files that can later call home and de-anonymise the attackers. Similar measures can be used against so-called 'anonymous' communication channels like Tor, given that the maximum latency of human communication is quite bounded.

In light of the power of strategic TA, there is little reason for Internet users to imagine that their online activities will remain private forever. Eventually, it must be assumed that disconnected communication segments will be reassembled into one complete stream. However, practically speaking, it is important to note that human communications are not random, and on many occasions they can even be predicted, either manually or automatically. With such intelligence in hand, eavesdroppers, criminals, and spies may already be lying in wait.

12 In many cases, attackers can hide behind fairly opaque IP ranges, but Whois is a good place for cyber defenders to start, and it is usually possible to lodge an abuse complaint to a Point of Contact listed here.

# 4. TRAFFIC ANALYSIS IN PRACTICE

TA is a strategic tool that can overcome many tactical defences, including encryption and covert channels. This is because cyberspace is a ubiquitous medium in which there are many ways that an eavesdropper can piece together otherwise obscure relationships and activities. At the very least, it is usually possible to detect that some type of communication is taking place, at which point an analyst can begin to isolate and evaluate specific data.

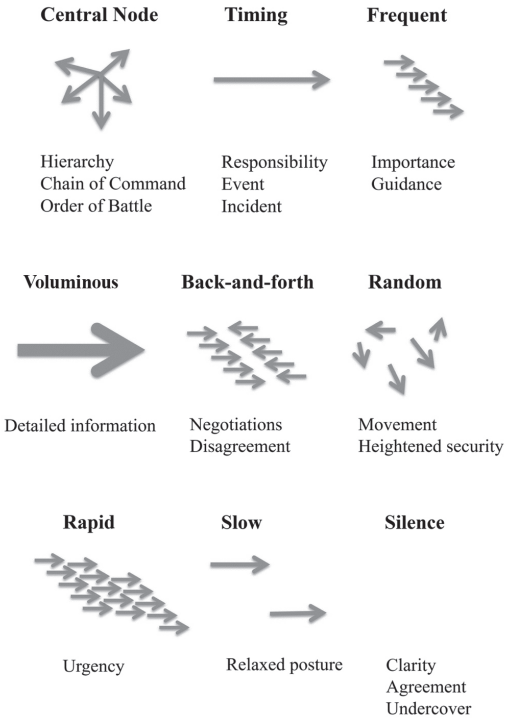
TA is both science and art. First, an analyst must collect sufficient data, from which they will develop a baseline for what appears to be normal traffic. Then the analyst seeks interesting patterns to examine further, often in the form of correlations and anomalies. In a sense, TA is analogous to speedreading, in that it provides a summary of a large volume of information from which the analyst can drill down and make intelligent discoveries.

Over time, TA researchers can solve even highly challenging problems, such as finding stealthy insiders, advanced cyber criminals and even Advanced Persistent Threat (APT) or nation-state actors. At this point, aggressive action may take place against an adversary, before the adversary is even aware that he or she has been discovered.

The list below (and in Figure 3) shows some of the ways in which communication patterns might be used to give away real-world activities.

- Central node
  - Hierarchy, chain of command, order of battle
- Timing
  - Responsibility for an event or incident
- Frequent
  - Importance or guidance
- Voluminous
  - Detailed information
- Back-and-forth
  - Negotiations or disagreement
- Random
  - Movement or heightened security
- Rapid
  - Urgency
- Slow
  - Relaxed posture
- Silence
  - Clarity, agreement, or undercover

**FIGURE 3.** TRAFFIC ANALYSIS: COMMON PATTERNS



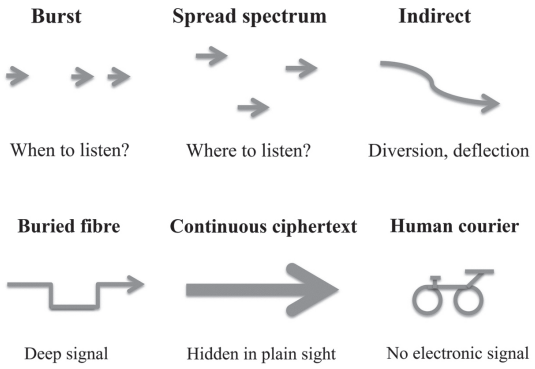

## *A. Counter-Traffic Analysis*

Historically, governments and militaries have defended against TA in many ways, especially by altering network traffic. The list below and Figure 4 summarise some of the most common methods.

- Bursts
  - So the attacker does not know when to listen
- Spread spectrum
  - So the attacker does not know where to listen
- Indirect routing
  - Diversion, deflection, appears intended for another party
- Buried fibre
  - Deep signal so the attacker cannot hear
- Continuous ciphertext
  - So the message is hidden in plain sight
- Human courier
  - So there is no electronic signal to capture.[13]

13 In 1998, a DARPA Challenge was issued, which recommended traffic padding and rerouting communications through long alternative network paths.

**FIGURE 4.** COUNTER-TA STRATEGIES AND TACTICS




The basic goal is to minimise exposure so that attackers cannot sense, monitor, analyse, jam, manipulate, or otherwise react to sensitive communications. However, high-level security can be impractical, as it is expensive, requires a determined effort to maintain and often attracts increased scrutiny. In general, only governments and large corporations can afford it.

On the Internet, there are many resources to defend against online attacks including TA: from Mixmaster, to Mixminion, Java Anon Proxy, and Tor, which obscure data such as email headers, web caches and network routing. With hackers, Snowden and Big Brother so often in the news, many researchers feel that developing digital security products, including to protect against TA, is a public good. For example, it is easy to understand that some levels of anonymity and privacy are needed for the proper functioning of democracy. That said, it is also clear that defence against strategic TA is difficult in the short run and may be impossible eventually.

## *B. Ethics and TA*

In the age of digital society and electronic government, the future of TA has profound implications for the world. Anyone with access to network logs has the capability to perform TA; however, governments, including law enforcement and intelligence agencies,[14] possess the capability to conduct strategic TA at will. Governments typically have 'backdoor' relationships with telecommunications providers, and businesses are mandated to retain logs far longer than typical business needs require.[15] Finally, commercial firms and advertising agencies have either collected or purchased granular TA for targeted purposes including tailored advertising.

At the international level, TA is fraught. Technology always outpaces policy and law. However, between nation-states, legal harmonisation is especially difficult in an environment where investigations impinge on another state's sovereignty and governments have become addicted to cyber espionage. The concern for privacy, democracy, and human rights is not uniform across the planet. When does routine surveillance become Big Brother? There is no doubt that decentralisation, privatisation and the fragmentation of telecommunications, including the

14 Articulating a rationale is easy: crime, espionage, and terrorism.

15 In the US, businesses have been required to maintain logs since September 2007 and Internet Service Providers (ISP) since March 2009.

creation of innumerable online web services that incorporate technologies such as Voice over Internet Protocol (VoIP), have complicated law enforcement and its jurisdictions; however, it is equally true that many governments have a vested interest in knowing too much about their citizens' private lives.

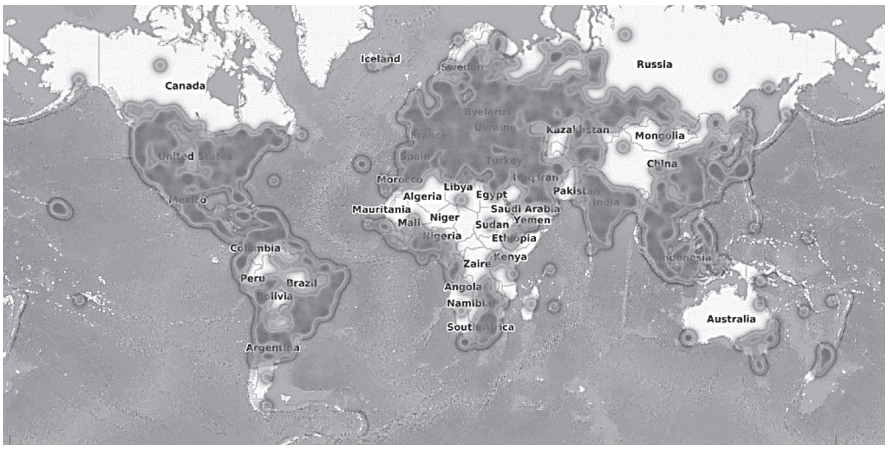
# 5. DATA ANALYSIS

In this section, the author will demonstrate some basic TA concepts against a large malware dataset, in order to show how quickly any analyst – for political, military, business, or personal reasons – can deduce an endless variety of real-world activities from computer log files.

Several layers of malware analysis will be presented, from the strategic to the tactical level.

Figure 5 displays a heat map of nearly a billion rows of malware data from the first six weeks of 2017.[16] In this dataset, nearly all 253-country code top-level domains (ccTLD) are represented.[17] The dark areas indicate where the author's firm detected malware; the white portions of the map have little malware due to their light concentrations of human settlement.

**FIGURE 5.** WORLD MAP OF DETECTED MALWARE, EARLY 2017

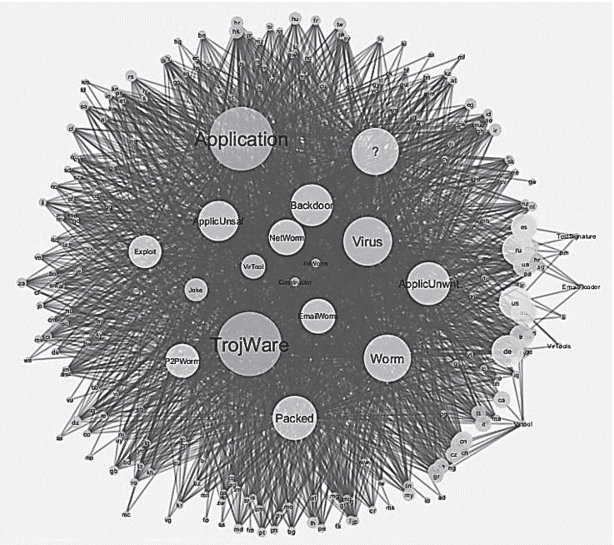



However, the heatmap is just a precursor to TA, which is more about connections and relationships. So, let us take a different view of the data. In Figure 6, the heatmap has been turned into a network chart, which shows a more logical relationship between malware categories and nation-states.

16 The author's cybersecurity firm analyses over 10 million potential pieces of malware every day, manages over 85 million cybersecurity software installations and works with hundreds of thousands of business customers and global partners around the world.

17 There are currently 250 ccTLDs, which are Internet top-level domains reserved for countries, sovereign states, or dependent territories. All ASCII ccTLD identifiers are two letters long, and all two-letter top-level domains are ccTLDs.

**FIGURE 6.** NETWORK MAP: MALWARE TO COUNTRY




Here, we can see that cyberspace really is to some degree a borderless domain in which attackers and malicious code move seamlessly between administrative and legal jurisdictions. The large circles represent malware categories, such as viruses and worms, while the smaller circles are the affected nation-states, identified by their ccTLDs.

TA is all about context. So, let us look at a more detailed description of the malware represented in Figure 6. In Figure 7, below, we can see that different types of malware affect different types of targets. The countries are ranked according to the ratio by which the 'backdoor' and 'worm' malware categories affected them in the first six weeks of 2017.

**FIGURE 7.** MALWARE COMPARISON: BACKDOORS VS. WORMS

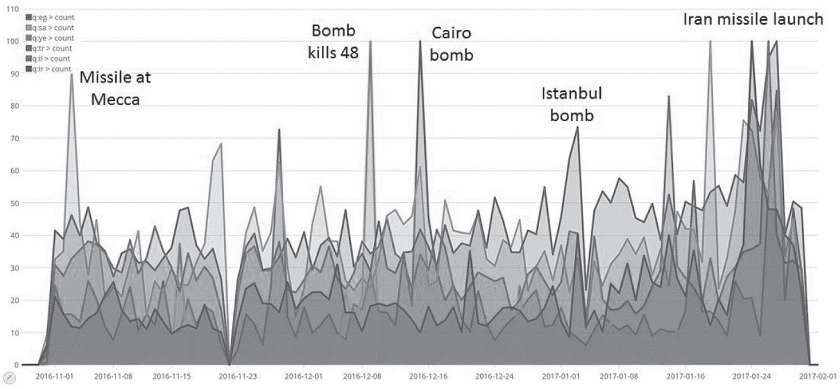
| Country | Backdoor Ratio Rank | GDP Rank (IMF) |
|---|---|---|
| Kuwait | 1 | 6 |
| Belgium | 2 | 25 |
| UAE | 3 | 9 |
| Portugal | 4 | 47 |
| Spain | 5 | 34 |
| Hong Kong | 6 | 12 |
| Iceland | 7 | 18 |
| Singapore | 8 | 4 |
| Bahrain | 9 | 17 |
| Madagascar | 10 | 179 |

| Country | Worm Ratio Rank | GDP Rank (IMF) |
|---|---|---|
| Nigeria | 1 | 131 |
| Ethiopia | 2 | 166 |
| Congo | 3 | 124 |
| Somalia | 4 | N/A |
| Maldives | 5 | 84 |
| Rwanda | 6 | 169 |
| Philippines | 7 | 120 |
| Bangladesh | 8 | 141 |
| Yemen | 9 | 158 |
| Moldova | 10 | 135 |

In Figure 7, we see that backdoors tend to afflict richer countries, which may be better protected against random attacks, and require more targeted strategies and tactics on the part of the attacker. The latter category, worms, afflicted poorer socioeconomic countries, which may run older, outdated, and therefore unsupported software, and are vulnerable to random digital attacks.

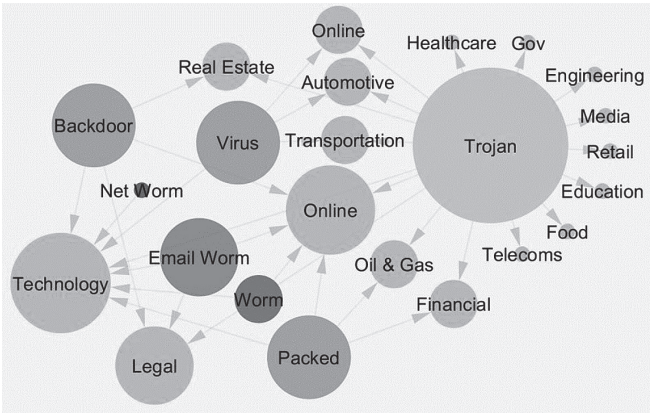
Time is the best friend of a TA researcher. Human conversations are necessarily confined within the boundaries of human patience. Figure 8, below, displays the overall number of malware incidents the author's firm detected in Egypt, Saudi Arabia, Yemen, Turkey, Israel and Iran during the first six weeks of 2017. Above the highest spikes for each country, the author has placed a key event from its national news, which took place on or about the same day. The malware incidents and geopolitical events are not necessarily related, but they may indicate law enforcement, intelligence, hacktivist, or criminal actions in cyberspace that may be correlative.

**FIGURE 8.** TIMELINE OF MALWARE ACTIVITY WITH GEOPOLITICAL OVERLAY



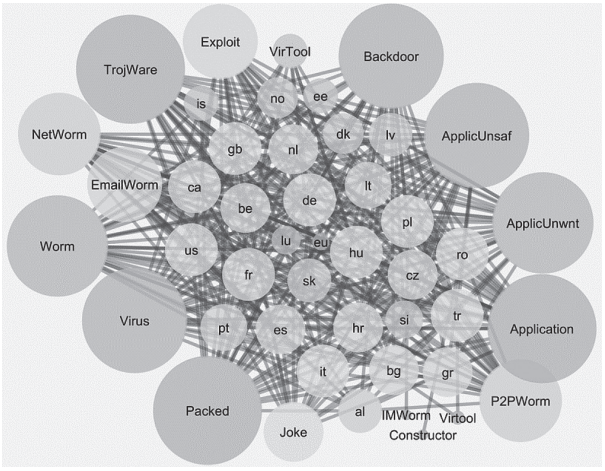

With TA, it is important to apply as much real-world logic as possible to an otherwise opaque dataset. As seen above, malware authors tailor code to a variety of targets, and not every economic sector, or 'vertical', is affected by every type of malware. In Figure 9, the author has paired his firm's top malware types against their target verticals, which highlight two quick conclusions: 1) trojans can be found in every vertical, and 2) attackers are using every type of malware to target the technology sector. This latter conclusion is not surprising, as the technology sector holds the keys to the virtual kingdom of cyberspace; in other words, compromising a particular software, website or protocol can lead to the compromise of all who use it.

**FIGURE 9.** NETWORK CHART: MALWARE TYPES VS. VERTICALS




Since this paper is written for the NATO CCD COE International Conference on Cyber Conflict (CyCon), it is useful to go back to the malware-to-country network chart in Figure 6, and delete the non-NATO countries (see Figure 10, below). A TA researcher can analyse the entire world (extrapolating from a sample dataset), or choose any part of it for dissection. Below, we can see at a glance that all NATO countries suffer from nearly all categories of malware.

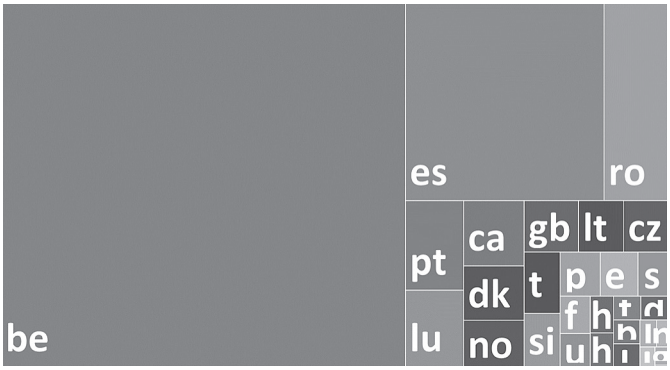
**FIGURE 10.** MALWARE MAP: NATO COUNTRIES




However, while the data in this network chart has been greatly simplified, it is still not granular enough. Therefore, I created an index for all of the NATO countries, based on the prevalence (by ratio) of each type of malware. In general, the malware-to-country pairings had somewhat similar profiles.

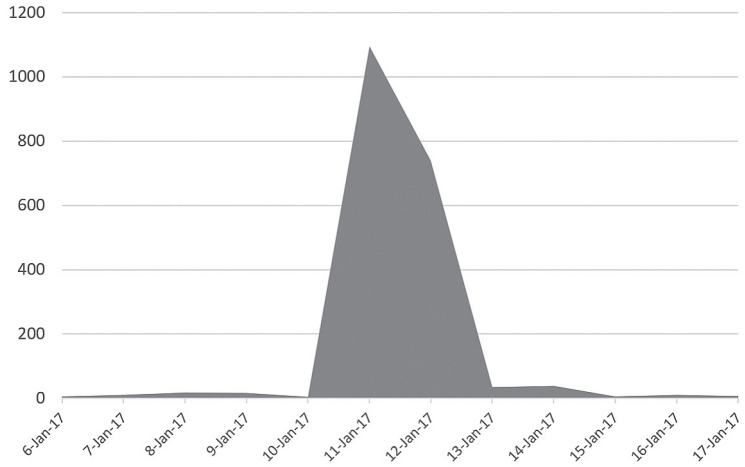
However, close TA almost always yields anomalies to investigate further. In this case, Belgium, which had a malware ratio that ranked near the bottom in nearly every malware category, scored a surprising first place in the 'backdoor' category. Figure 11, below, clearly shows Belgium's domination of this subset of the malware data.

**FIGURE 11.** BACKDOOR DETECTIONS IN NATO COUNTRIES: EARLY 2017




Next, we should find out precisely when the backdoor detections took place. The timeline in Figure 12 shows this quite clearly: 10-13 January 2017. At this point, we do not know what caused the sudden increase in backdoor detections. It is possible that new security signatures simply found older, previously installed malware. It is also possible that there was a targeted attack against one or more enterprises in Belgium, and that the backdoors were installed with the aid of phishing, a worm, social engineering, or the unwise installation of a malicious application.

**FIGURE 12.** TIMELINE: BACKDOOR DETECTION IN BELGIUM



At this stage, TA has successfully taken the analyst from a strategic to a tactical level. However, at this point, there are still a variety of ways for a researcher or an investigator to move forward. For example, this information could simply be given to network system administrators for device isolation, digital forensics, software patching, reinstallation and further attack mitigation. Alternatively, investigators may keep this digital TA quiet (leaving the backdoors open temporarily), and begin a process of real-world correlation, in the hope of ensnaring a sophisticated adversary. Such an investigation would include asking hard questions such as: What else happened during this time period that may shed light on the malware's ultimate purpose? If the victim was a business, were there any important trade deals happening at the time? If the victim was a government, did the sharp spike occur just prior to an election, or a national security incident?

Digital TA can inspect log files for any kind of correlation, of a political, military, criminal, business, or personal nature. With enough data in hand, sometimes gathered over many years of painstaking intelligence collection, TA can even help to solve the attribution problem, or that of the anonymous hacker. There are many prominent historical examples, including The Cuckoo's Egg, Moonlight Maze, Stuxnet, Sony, and the Democratic National Committee (DNC).

In a geopolitical context, no stone will be left unturned. TA will not only encompass the temporal, spatial, directional, and logical character of network traffic, but will incorporate intelligence from other domains as well, such as human intelligence (HUMINT), signals intelligence (SIGINT), and open source intelligence (OSINT).

# 6. CONCLUSION

Historically, TA has been used to circumvent a wide range of core security protocols including encryption. Businesses use TA for market research and advertising; governments collect foreign and domestic intelligence; researchers analyse countless streams of data for academic papers. As we grow more dependent on the Internet – and give IP addresses to everything from toasters to the brakes on our cars – the power of IT will strengthen, magnify, and amplify TA as never before.

This paper has sought to bring TA, especially its digital version, to a wider audience, by describing not only its famous achievements during World War II, but how modern computer log files are essentially a record of all human activity, and can be mined for virtually any kind of intelligence value, from the strategic to the tactical level. Because cyberspace is merely a reflection of human affairs, all major geopolitical events, from elections to invasions, have digital analogues that are just waiting to be discovered.

TA has limitations. Computer log files can provide convincing evidence of real world activity, but once the analysis is complete, traditional investigative practices, such as physical (and network) forensics must complement TA. For example, even the most famous cyberattack case studies, from Moonlight Maze to Estonia and Sony to the DNC, have remained mired in the

'attribution' controversy for years, despite the fact that many analysts believe a preponderance of evidence, including TA, points to a guilty party. At the end of the day, TA is just one tool in a larger toolbox, but it can serve to complement other, more traditional, tools in striking ways.

Looking forward, TA is an established discipline, but research gaps will continue to remain due to the rapid and dynamic evolution of all things IT. Future research should investigate the effect of cloud computing, autonomous systems, artificial intelligence (AI) and more. More analysis of the legal and ethical aspects of TA, especially given that there is only one Internet and one cyberspace which encompasses every jurisdiction and sovereignty on Earth, is also long overdue.

## REFERENCES


[1] M. Fathy and M. Y. Siyal. 'An image detection technique based on morphological edge detection and background differencing for real-time traffic analysis.' *Pattern Recogn. Lett*. 16, 12, 1321-1330. December 1995.

[2] L. Wischhof, A. Ebner, and H. Rohling. 'Self-organizing traffic information system based on car-to-car communication: Prototype implementation.' *International Workshop on Intelligent Transportation (WIT)*. Hamburg. March 2004.

[3] Chevalier and L. M. Wein. 'Scheduling Networks of Queues: Heavy Traffic Analysis of a Multistation Closed Network.' *Operations Research* 41(4), August 1993.

[4] J.-F. Raymond. 'Traffic Analysis: Protocols, Attacks, Design Issues and Open Problems.' *Zero-Knowledge Systems*, December 19, 2000.

[5] G. Varatkar and R. Marculescu. 'Traffic Analysis for On-chip Networks Design of Multimedia Applications.' *DAC 2002*, New Orleans, Louisiana, USA, June 10-14, 2002.

[6] W. Lee, C. Wang, and D. Dagon (Eds). *Botnet Detection: Countering the Largest Security Threat*. New York: Springer, 2008.

[7] C. Mazzariello. 'IRC Traffic Analysis for Botnet Detection.' *The Fourth International Conference on Information Assurance and Security*, IEEE, 978-0-7695-3324-7/08. 2008.

[8] G. Danezis. 'The Traffic Analysis of Continuous-Time Mixes.' In: D. Martin, A. Serjantov (eds) *Privacy Enhancing Technologies*. Lecture Notes in Computer Science, 3424. Springer, Berlin, Heidelberg. PET 2004.

[9] M. Crovella and E. Kolaczyk. 'Graph Wavelets for Spatial Traffic Analysis.' BUCS-TR-2002-020, Office of Naval Research. July 15, 2002.

[10] G. Danezis and R. Clayton. 'Introducing Traffic Analysis.' Chapter in *Digital Privacy*. Auerbach Publications, ISBN: 9781420052176. January 21, 2007.

[11] N. Mathewson, R. Dingledine. 'Practical Traffic Analysis: Extending and Resisting Statistical Disclosure.' In: D. Martin, A. Serjantov (eds) *Privacy Enhancing Technologies*. Lecture Notes in Computer Science, 3424. Springer, Berlin, Heidelberg. PET 2004.

[12] M.-Sup Kim, Y. J. Won, and J. W. Hong. 'Characteristic analysis of internet traffic from the perspective of flows, Computer Communications.' Vol. 29, Issue 10, 1639–1652, *Monitoring and Measurements of IP Networks*. June 19, 2006.

[13] J. R. Goodall, W. G. Lutters, Rheingans, and A. Komlodi. 'Focusing on Context in Network Traffic Analysis.' *Visualization for Cybersecurity*, 0272-1716/06, IEEE Computer Society. March/April 2006.

[14] J. Deng, R. Han, and S. Mishra. 'Countermeasures Against Traffic Analysis Attacks in Wireless Sensor Networks.' *Technical Report CU-CS-987-04*. December 2004.

[15] J. Deng, R. Han, and S. Mishra. 'Decorrelating Wireless Sensor Network Traffic To Inhibit Traffic Analysis Attacks.' *Elsevier Pervasive and Mobile Computing Journal*, Special Issue on Security in Wireless Mobile Computing Systems, 2, issue 2, 159-186. April 2006.

[16] J. Deng, R. Han, and S. Mishra. 'Intrusion tolerance strategies in wireless sensor networks.' In *Proc. of IEEE 2004 International Conference on Dependable Systems and Networks (DSN'04)*. 2004.

[17] C.V. Wright, S.E. Coull, F. Monrose. 'Traffic morphing: an efficient defense against statistical traffic analysis.' In: *Proc. of ISOC Network and Distributed System Security Symposium (NDSS)*. 2009.

[18] N. Mathewson and R. Dingledine. 'Practical traffic analysis: Extending and resisting statistical disclosure.' In *Proceedings of Privacy Enhancing Technologies workshop (PET 2004)*, LNCS. May 2004.
[19] D. L. Chaum. 'Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms.' *Communications*, 24 ACM Number 2 0001-0782/81/0200-0084. February 1981.
[20] R. E. Newman, I. S. Moskowitz, Syverson, and A. Serjantov. *Metrics for Traffic Analysis Prevention*. Office of Naval Research. 2003.
[21] M. Qadeer, A. Iqbal, M. Zahid, and M. Siddiqui. 'Network Traffic Analysis and Intrusion Detection using Packet Sniffer.' *Second International Conference on Communication Software and Networks*, 978-0-7695-3961-4/10 IEEE. 2010.
[22] K. Dyer, S. E. Coull, T. Ristenpart, and T. Shrimpton. 'Peek-a-Boo, I Still See You: Why Efficient Traffic Analysis Countermeasures Fail.' *IEEE Symposium on Security and Privacy*. 2012.
[23] X. Gong, N. Kiyavash, and N. Borisov. 'Fingerprinting Websites Using Remote Traffic Analysis.' *CCS'10*, Chicago, Illinois, USA. ACM 978-1-4503-0244-9/10/10. October 4-8, 2010.
[24] S. Sarvotham, R. Riedi, and R. Baraniuk. 'Connection-level Analysis and Modeling of Network Traffic.' *IMW'OI*. San Francisco, CA, USA, ACM l-581 13-435-5. November l-2, 2001.
[25] N. Ben Azzouna and F. Guillemin. 'Analysis of ADSL traffic on an IP backbone link.' *France Telecom R&D*, 0-7803-7975-6/03. 2003.
[26] J. Schonwalder, A. Pras, M. Harvan, J. Schippers, and R. van de Meent. 'SNMP Traffic Analysis: Approaches, Tools, and First Results.' 1-4244-0799-0/07 *IEEE*. 2007.
[27] T. Silverston, O. Fourmaux, A. Botta, A. Dainotti, A. Pescapé, G. Ventre, and K. Salamatian. 'Traffic analysis of peer-to-peer IPTV communities.' *Computer Networks*. Elsevier. 2008.
[28] M.-Sup Kim, H.-Jung Kang, and J. W. Hong. 'Towards Peer-to-Peer Traffic Analysis Using Flows.' In: Brunner M., Keller A. (eds) *Self-Managing Distributed Systems*. Lecture Notes in Computer Science, 2867. Springer, Berlin, Heidelberg. DSOM 2003.
[29] C. Rossow, C. Dietrich, H. Bos, L. Cavallaro, M. van Steen, F. C. Freiling, and N. Pohlmann. 'Sandnet: Network Traffic Analysis of Malicious Software.' *BADGERS 2011 Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, ACM. Salzburg. 2011.
[30] S. J. Murdoch and Zielinski. 'Sampled Traffic Analysis by Internet-Exchange-Level Adversaries.' In *Privacy Enhancing Technologies (PET)*. 2007.
[31] J. Kinicki and M. Claypool. 'Traffic Analysis of Avatars in Second Life.' *NOSSDAV '08*, ACM 978-1-60588-157-6/05/2008. 2008.
[32] S. Fernandes, R. Antonello, J. Moreira, D. Sadok, and C. Kamienski. 'Traffic Analysis Beyond This World: The Case of Second Life.' *NOSSDAV'07*, Urbana, Illinois USA, ACM 978-1-59593-746-9/06/2007. 2007.
[33] Svoboda, W. Karner, M. Rupp. 'Traffic Analysis and Modeling for World of Warcraft.' *ICC 2007 proceedings of the IEEE Communications Society*, 1-4244-0353-7/07. 2007.
[34] M. Suznjevic, M. Matijasevic, and O. Dobrijevic. 'Action specific Massive Multiplayer Online Role Playing Games traffic analysis: Case study of World of Warcraft.' *NetGames '08*, 2008 ACM 978-160558-132-3-10/21/2008. 2008.
[35] W.-chang Feng, F. Chang, W.u-chi Feng, and J. Walpole. 'Provisioning On-line Games: A Traffic Analysis of a Busy Counter-Strike Server.' *OGI CSE Technical Report*, CSE-02-005, Portland State University. May 15, 2002.
[36] K.-Ta Chen, Huang, and C.-Laung Lei. 'Game traffic analysis: An MMORPG perspective.' 1389-1286, *Computer Networks* 50 3002-3023 Elsevier. 2006.
[37] K.-Ta Chen, J.-Wei Jiang, Huang, H.-Hua Chu, C.-Laung Lei, and W.-Chin Chen. 'Identifying MMORPG Bots: A Traffic Analysis Approach.' *EURASIP Journal on Advances in Signal Processing*. 2009.
[38] A. Back, U. Moller, and A. Stiglic. 'Traffic analysis attacks and trade-offs in anonymity providing systems.' In I.S. Moskowitz, editor, *Information Hiding*, pp. 245-257. Springer-Verlag, LNCS 2137. 2001.
[39] S. J. Murdoch and G. Danezis. 'Low-Cost Traffic Analysis of Tor.' 2005 *IEEE Symposium on Security and Privacy*, Oakland, California, USA. May8-11, 2005.
[40] F. Fusco and D. Luca. 'High Speed Network Traffic Analysis with Commodity Multi-core Systems.' IMC'10, Melbourne, Australia, ACM 978-1-4503-0057-5/10/11. November 1-3, 2010.
[41] Y. Lee, W. Kang, and H. Son. 'An Internet Traffic Analysis Method with MapReduce.' *2010 IEEE/IFIP Network Operations and Management Symposium Workshops* 978-1-4244-6039-7/10. 2010.
[42] L. Xiao, J. Gerth, and Hanrahan. 'Enhancing visual analysis of network traffic using a knowledge representation.' *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, pp. 107-114. 2006.
[43] J. R. Goodall, W. G. Lutters, Rheingans, and A. Komlodi. 'Preserving the Big Picture: Visual Network Traffic Analysis with TNV.' *Workshop on Visualization for Computer Security*, 0-7803-9477-1/05 IEEE. 2005.