doi:10.2498/cit.1002282

The Dark Web: Cyber-Security Intelligence Gathering Opportunities, Risks and Rewards

Gregory Epiphaniou, Tim French and Carsten Maple

Cybersecurity Research Group (CyBSeG), IRAC, Department of Computer Science and Technology, University of Bedfordshire, Luton, United Kingdom

We offer a partial articulation of the threats and opportunities posed by the so-called Dark Web (DW). We go on to propose a novel DW attack detection and prediction model. Signalling aspects are considered wherein the DW is seen to comprise a low cost signaling environment. This holds inherent dangers as well as rewards for investigators as well as those with criminal intent. Suspected DW perpetrators typically act entirely in their own self-interest (e.g. illicit financial gain, terrorism, propagation of extremist views, extreme forms of racism, pornography, and politics; so-called 'radicalisation'). DW investigators therefore need to be suitably risk aware such that the construction of a credible legally admissible, robust evidence trail does not expose investigators to undue operational or legal risk.

Keywords: dark web, predictive model, signalling

1. Introduction

The so-called "Hidden" or Dark Web (HW;DW) comprises a heterogeneous collection of P2P oriented Intranet/Internet Open Source customisable search engines (such as 'YaCy' `http://yacy.net`), "*Search by the People for the People*"; distributed and dynamic P2P file-exchange e-communities that in turn support an equally diverse, "rich" collection of DW accessible digital materials. DW resources are not always fully visible to generic search engines such as Google, Mozilla, Firefox or Opera. Rather, DW sites (including "black-listed" DNS) are typically only visible, hence exploitable both by legitimate users and those with criminal/illicit intentions through the use of a variety of Open-source (i.e. community authored) tools that leverage, hence seek to exploit various "niche" DW e-silos. This presents a technical challenge wherein an e-silo is physically distributed (i.e. fragmented across thousands of nodes known as P2P clients). Typically, freeware downloads provide client access for nontechnical users. Devices can be leveraged by the DW with or without a user's explicit consent or knowledge. Furthermore, identity "hiding" (e.g. IP address obfuscation) is integral to the DW. Public domain tools are freely available that quickly delete (beyond forensic recovery) both client-side files and DW related user browsing/privacy traces. However, recently [1], there has been some progress as regards the capture of client side registry "traces" that can establish previous uTorrent (i.e. P2P) usage, usage patterns unknown to the user of the uTorrent client. Users of the HW may therefore be engaging in activities under a false premise: namely, their activities are "untraceable", when in reality, they may be exposing themselves to unquantifiable risks. Of course, such HW registry "traces" are not typically robust. However, the main drawback of this approach is obvious: we often do not know, let alone establish with certainty, in a timely and opportune manner, their location, hence location of local client devices, remote device storage facilities or otherwise seize suspect devices. Human goal and usage diversity with partial investigative opacity are central to the *modus operandi* of the DW. From a theoretical perspective, Gambetta [2] has previously explored various ways in which mimicry, deception and identity verification operate within

The first and abridged version of this paper appeared originally in LISS 2013, Reading, UK, August 21–24, 2013. Published in partnership with ICIR of Beijing Jiaotong University.

criminal gangs such as the Mafia. His perspective is sociological rather than linguistic, informally presented rather than rigorous [2]. Criminals typically seek to communicate in ways designed to obfuscate their "true" identities. This raises intriguing notions of mimicry and deception in a so-called semi-sorted equilibrium signalling environment wherein it is relatively cheap to emit false surface level signs as between signaller-receiver in which "false" identity cues are emitted by an individual or group of individuals within a given community of interest. This serves to encourage false inference formation as to the (inferable but ultimately unknowable) hidden cognitive trust-warranting properties (such as honesty, integrity, unselfishness, reliability). The true motives hence intents of an active participant within a given DW community of interest (*consumer, receiver, uploader, tracker, content moderator*) may be partially or fully opaque to those who are not members of that community of interest. This serves to encourage mimics (using false *manifesta*), since it is relatively easy i.e. low cost, to create a false set of e-identities or set of e-personas. Typically, signallers use a restricted code of some kind, wherein a linguistic cue, code word, series of words, phrases derived often from normative interlocutory acts (i.e. apparently "innocent" speech acts) serve to obfuscate and plan criminal intents and actions [3].

2. Signalling Aspects: Low Cost, High Risk?

One of the relatively unexplored areas from an academic perspective is that of the signalling exchanges between sender and recipient within Torrent DW contexts. Though Gambetta [1] has sought to partially explicate signalling within high cost criminal closed worlds such as prison environments, and French [3] has developed a semi-formalisation of signalling within e-commerce contexts wherein mimicry and deception are often manifest, little work has hitherto been undertaken to formalise DW e-silo signallers within the specific context of a Torrent VO community of interest. Table 1 (below) seeks to offer an enumeration of some of the most obvious signaller/receiver node types that also seeks to characterise victimisation aspects and maps these to levels of an e-trust ladder given elsewhere [3]. Paradoxically, within low-cost signalling contexts such as e-DW communities of interest, it is often relatively simple for anyone to join or to receive illegal (e.g. "*extreme*" pornographic materials) using freely available user-friendly/installable uTorrent clients. However, this ease of adoption and accessibility does confer dangers with respect to victimisation and incrimination aspects of those with "honest" intent, who are seeking to engage (as mimic participants) for the purposes of gathering evidence from within DW VO communities.

Given that cyber-traffic analysis across a multitude of nodes wherein payloads are distributed across hundreds (theoretically thousands) of P2P DW participants is in itself a complex, not to say potentially intractable area, the Table 1 represents an idealistic reverse engineered set of participants. Despite difficulties for investigators, it is feasible to not only identify the main players (ie. active nodes), but also to analyse cyber-traffic patterns between nodes.

Role	**Relative Cost**	**Danger of Victimisation?**	**e-Trust level[3]**
DW Member (receiver only)	Low	High	Social, semantic, syntactic
DW Member (sender/receiver) e.g. "moderator"	Low	Low	Social, semantic, syntactic+empiric
DW Victim (e.g. photographed, coerced or "willing") participant	Medium/high	V High	Social, semantic
Casual visitor/surfer/accidental visitor (receiver only)	Low	High	Semantic, syntactic
Covert Police/law enforcement operative (Trojan)/Overt participant mimic	Low	Low	Social, semantic, syntactic

Table 1. DW VO signal and receivers: extreme pornography scenario.

2.1. A Semiotic-theoretic Approach to Low Cost Signalling

Computational models of trust mechanisms based on explicating notion of trust in the context of VO e-services have only recently emerged [4]. One need for this is that traditional security mechanisms are being increasingly challenged by open, large scale and decentralised environments. This situation naturally leads to a highly decentralised model of security, risk and trust between VO partners in some pre-determined orchestrated manner. Several works are currently examining relevant trust issues at the VO level of abstraction, including the work from the TrustCOM project [4]. These works claim to deal with high-level "reputation" issues. However, much of these works actually seek merely to address tangible security aspects and performance aspects. Among the first works to establish the need to examine "soft" trust issues are those described in Song [5]. Their trust index is calculated using a mixture of inputs including the site's defence capabilities and site reputation, defined as a performance track record. Their solution is relatively "heavy-weight". A large number of inputs are used to calculate the trust index, via the use of neural network based techniques.

Nevertheless, many kinds of deception on the web involve low cost identity signalling which engender false trust semiosis due to low cost of emitting false signals. One of the most common is so-called 'phishing' attacks. These have been considered in some detail by [6]. In essence, a phishing attack is a form of deception in which a consumer is tricked into logging on to a fake website by means of a fake e-mail in order to steal their identity. It is successful because of the cheapness with which an adversary can emit false identity signs (i.e. create an e-mail containing an embedded URL, that lures the online consumer into the trap). One reason why such deceptions are effective is the conceptual gap between a typical e-banking site customer's model of the IT system and the IT system's actual behaviour.

Another reason is that of a temporal delay as regards semiosis. Sign and signal exchanges at the protocol level are ultimately interpreted by human agents. Such surface level signs usually take the form of system messages, rendered using a browser, an e-mail client or similar tool. However, there exists a temporal delay, as well as an information gap and trust asymmetry, whereby the human actors are only able to correctly interpret the meanings of these exchanges (infer trust or mistrust) after a lengthy and 'hidden' sign and signal exchange between 'hidden' or partially visible technological agents. Often such agents generate messages at the HCI level that users cannot easily interpret (such as Digital Certificate "out of date"). Some signs (hence signals) are more reliable than others, hence decoding such signs and assigning meaning in the form of trust signals is a game beset by deception and mimicry. Below the interface level at the machine level, protocols such as SSL/TLS rely on the 'hidden' exchange of signals between server and client. These signals are not directly observable by humans at the time of exchange, but are later presented to users in a delayed manner in the form of system messages and warnings by the software.

Trust has been studied from each and every angle: in the philosophical, sociological, psychological, scientific, economic, and legal sense – to name just a few. One perennial barrier to synthesising a definitive trust model and theory of trust is the lack of agreement as to definitions of trust. Indeed, one of the central difficulties is that the notion of trust is closely related to other concepts such as ability, benevolence, integrity, reliance, competence, and credibility [7]. Nevertheless, it is possible to identify a core body of work with specific context to trust and usability [8]. Cultural factors pre-determine a consumer's given set of trust expectations (trust branding, text vs. graphics, written narrative and on-line guarantees), [9]. Further, that these expectations match generic societal and social attitudes and social structures within target countries. The premise is that a set of given computer based signs assembled into a coherent user experience induce differential consumer reactions according to a particular consumers cultural orientation and belief system. As yet, however, the field of cross-cultural trust research is relatively immature. For example, it is still not known with any certainty, how trust formation differs between one target group and another, other than by exhaustive empirical and comparative studies. The evidence that has emerged is at best tentative in relation to trust issues.

Indeed, numerous studies have identified a number of trust building factors within the on-line user experience. These include: an effective navigation model, contact information, the embedding of human forms, and the use of trust seals. These and other design features combine so as to create what might best be termed a measurable user confidence level, mediated by inherent risk and reward. This confidence level appears to comprise both an affective and rational component [4]. Hence, unintended or rather general system interface properties such as the use of colour combinations, visual metaphors, or the use of specific types of fonts (etc.) can potentially influence user trust formation – not just obvious tangible security features or trust seals. In essence, it seems that various surface level signs and signals of trust are perceived (decoded) by users in relation to trust aspects of an interface. From these signs and signals, users infer hidden trust warranting properties such as benevolence, honesty, and integrity (etc.).

Furthermore, users build an expectation as to the future behavioural properties of an interface as they engage. Thus, we trust a user interface to perform a particular task if and only if we expect that the likelihood that the system will fulfil its obligation lies above our own personal trust confidence threshold. This notion has important implications. Namely, both honest and dishonest (malevolent) DW designers should not simply assume that specific trust building factors or individual trust thresholds are universal across every culture, individual or usability context. Indeed, there is ample evidence to the contrary [10]. Nor should they assume that confidence levels and behavioural expectations remain static: rather, they are dynamic and are continuously informed and driven by external events in the "real world" of the user and also, of course, by the user's own immediate and direct experience of the interface itself. (One only has to think of the recent "credit crunch" and the consequent measurable global rise of distrust in banks to realise how easily external events can dynamically influence trust thresholds). Trust building factors and trust requirements need therefore to be explicitly addressed as an integral part of interface design. If usability trust requirements are simply taken for granted, or allowed to remain implicit, then user adoption and acceptance may be compromised.

Recommended Actions: Trust requirements (both intangible and tangible security) should be separately articulated, but need to be seen as integral to the specification, analysis, and design, of a computer interface. System designers should make such trust requirements fully explicit. It is also necessary to consider the role of rational judgement as well as the role of emotion (affect). Do not make the mistake of thinking that embedding visible tangible security signs alone equates to trust building. It is more likely that users decode a particular interface design as they experience it, using a wide variety of interface features so as to build their own confidence level. Try therefore to seek ways to support user trust formation – by designing "in" known trust building properties that not only support user confidence levels, but which also reinforce positive user expectations. For honest designers it is essential to identify signs of trust that are not easily reproducible by malevolent designers (e.g. designers of fake anti-virus software packages that actually serve to infect rather than disinfect client devices). Indeed, the latter area is worthy of a separate research initiative in its own right, that is, deployment of trust signs within an environment beset by mimicry.

3. Dark Web Groups, Attack *Modus Operandi* and Harvesters

The DW offers a virtual-world wherein information based cyber-warfare is being carried out between organised crime, terrorist groups, extremists of various kinds and the agencies acting on behalf of a civil society. There has been limited focused crawling work carried out in relation to the DW. Prior hidden web research has sought to mostly focus upon automated form filling or query redirection to hidden databases, i.e., *accessibility* issues. There has been little emphasis on building topic-specific web page collections from these hidden sources: that is to say, specifically targeted at harvesting Dark Web content pertaining to hate and extremist groups. Harvesting the DW is certainly non-trivial, however academics have entered this domain with some success [11]. Leveraging AI techniques that accurately identify the authorship of heterogeneous postings and web-site contents, Chen [11] and his co-workers were able to identify DW groups and trace their activities. As a low cost semi-sorted equilibrium environment, the DW offers a signalling environment which is inherently open, transparent and is beset by mimicry. This offers potential for deception and obfuscation; equally, it is possible through using high-performance computational resources to reverse engineer DW groups and their activities from an analysis of "big-data" i.e. DW traffic, postings, web-sites and their contents. However, it would appear that there exists an asymmetry wherein DW activities have remained largely undetected.

With respect to our present *foci*, DW groups of potential interest are those that launch active attacks against intellectual property that typically lies exposed to an internet connection. This comprises a vast data set, wherein patents, secrets, inventions are potentially targeted (harvested) by criminal gangs. Typically, such attacks have compromised large corporate servers, e-Government data silos and so forth. Some have coined these activities "crimeware" whilst for others the field is a broad one comprising many forms of simple to complex attack configurations, typically, but not exclusively orchestrated by DW groups. The so-called "Hidden Lynx" cyberspy gang has waged targeted attacks since at least 2009. Attacks include the injection of malware into legitimate websites frequented by their targeted sectors, mainly from financial services firms in the U.S. Symantec has alleged that the gang is also tied to Operation Aurora, which targeted Google, Intel, Adobe, and other major U.S. firms, which was revealed in 2010. The main concern from a national security perspective are those groups who remain partially or completely covert wherein the first that is known is reverse engineered from a "zero" day attack.

4. Attack Detection Using Evolving Adversarial Behaviour in Cyber Clusters

Cyber attack projection has been classified as an L3 fusion problem based on a revised model introduced in [12]. The added complexity from dynamic changing parameters has been studied in [13] where a separation between the cyber attack methods used for detection and the modelling of network configuration is achieved. In an attempt to assure untraceability and undetectability between entities, as part of the communication process, confidentiality mechanisms have proved insufficient to address attacks against those principles. In addition, different international laws and cross-continental cyber crimes are still far away from getting affected by a global public policy on allowed/prohibited data flows. The complexity and decentralisation management provided by anonymous networks can lead to information leakage that can be potentially catastrophic in the hands of an adversary. This plethora of virtual connections over multiple hops can offer the privacy required between the communication parties making identification and traceability a difficult task to be achieved with some trade offs, especially between anonymity and performance, in dynamic P2P establishments. Assuming that the notion of security in any electronic communication is not absolute, even with minimization of exploitable bugs in a given application, this does not assure that the number of attack vectors will be reduced, but rather change. The outcomes from each individual assessment are then fused so as to determine targeted entities (i.e. attack intentions). Contextual information from IDS logs is also used as an input in several studies using a variety of tools for that purpose [14][15]. The work presented at this section is an extension of our work presented in [16] where both the model and the resulting algorithm are formalised.

We present a preliminary P2P intrusion forecasting model based on a "*guilty by association*" approach. We employ the notion of a cyber community *Cp* where *p* is the number of participants (nodes) in any given "*transaction*", wherein $p \geq 2$. We do not attempt here to differentiate between different types of P2P communicative acts within or across shared DW *Cp* communities. We divide a cluster *s* of cyberspace *S* based on a set *A*, of parameters that represent a set of common interests between pairs of active nodes in a given cyber community C_p. The actual scale of such a communication network can play a crucial role in terms of the anonymity level preserved as part of its core operation(s). There is also a distinct trade off between the anonymity preservation properties of such networks and the size and number of participants in the actual communication process with elements such as trust and reputation adding further complexity to an attempt to quantify anonymity and privacy levels.

Each *p* in *Cp* typically manifests itself as a set of network traffic patterns *Tp*. Namely, *Tp* are specific to the requirements hence communication protocols used within any *Cp*. Identifying *Cp* is an "NP-hard" problem. This is due to the complexity of *Cp* possible node cluster configurations (sparse, loose, and richly connected), that comprise *Cp*. The space is compartmentalized and expanded exponentially as a set of *s* clusters such that members can join multiple

Cp, hence, engage in DW P2P activities. We define a cyber community of "shared interests" for a given cluster as:

$$C_p = \frac{\sum_{i=2}^{n} p_i * T_p}{S} \tag{1}$$

We postulate that the probability of each attacker for any given *Cp* can be expressed as:

$$p\left(\alpha_i^{T\rho}\right) \tag{2}$$

Although the communication protocol is common within transactions between nodes in a highly decentralised fashion, logical accessibility from one cluster to a different one is subject to different security requirements. These include vulnerabilities and dynamically changing attack vectors. A typical example is advanced invasive attacks where re-engineering of the attack vectors to evade security mechanisms at the network perimeter is a common manifestation of such multi-stage cyber attacks. The engagement participation status, changes over time, thus:

$$\sum_{i=1}^{T\rho} p\left(\alpha_i^{T_p}\right) = 1 \tag{3}$$

We propose that each participative communicative state-space path will provide a reliable estimate for the set of sender-receiver communicative temporal dynamics between potential attacker-to-victim pairs, within or across single or multiple *Cp*. Zero-trust is assumed in this environment as data and service metrics for each node define equal importance to other nodes in the same or different clusters *s*. This renders all nodes relevant and important to the criticality of a secure network operation. Attackers' motive might be diverse as focus can be on any node in the cluster, given that they all carry equal criticality for overall cluster health. Information from updated CVE databases can be used to quantify criticality levels in each cluster for different services provided (local or remote) for each node $p \subset s$. Each cluster could be subject to certain sequenced attacks that generate different alerts at each node *p* with at least 2 nodes presented in that cluster.

The nature of the protocol and network used does not seek to accommodate a clearly defined network perimeter where alerts are centrally generated. Our model assumes little information regarding prior attack behaviour or attack attributes in a given cluster. Failed attacks can also yield information about attack capabilities and motives within one or more clusters in a P2P network. For example, DHT routing poisoning attacks can yield valuable information about DoS capabilities an adversary might employ against a specific node or a set of nodes within a specific cluster. The problem of entity-to-identity mapping in such environments plays a crucial role also in identifying attackers' capabilities. If an adversary can leverage a number of multiple identities he/she may control a significant portion of the cluster, thus dictating adaptations to current threat modelling. Immediate affects to data integrity and redundancy have been manifested in such networks as a result of this mapping [17][18][19]. It is assumed that field intelligence can be used so as to increase predictive accuracy over time. Following the work of [20], specifically their Exponential Weighted Moving Average (EWMA), the participation rate over a given time period *t*, is expressed as:

$$\gamma_{a_1 c_p}^{T\rho}(t{+}1) = \frac{\prod_{t_1}^{t} \beta(1-\beta)^{t-(t-1)} * \gamma_{a_1 c_p}^{T_p}(t') * C_p}{100} \tag{4}$$

where β is the smoothing coefficient; t' the span of the attacker's participation window presented to *Cp*, and 1 being the first time an attacker can be linked with any given community, with $(t{+}1)$ being the most recent. We assume the probability of an attack increases as the association rate of an attacker increases. It follows that:

$$\gamma_{a_i c_p}^{T_p}(t+1) > 1 \tag{5}$$

This serves to increase attack probability. Participation alone, without the evidence derived from inter-nodal activities, is not sufficient. The population density of a cluster plays a strong protagonist role in providing an accurate estimate of future attacks. The transaction rate (T_r), being a function of cluster density:

$$T_r = \frac{\left[\sum_{i=1}^{N} T\rho\right] * \gamma_{a_i c_p}^{T_p}}{100} \tag{6}$$

As the infrastructure is a P2P environment, there is no distinction to be made between internal and external network segments. The assumption is made that each cluster s is not seen in isolation from the rest of the space S as each node p can simultaneously participate in more than one s at the same time. Indeed, deriving models that capture evolving attacker behaviour is far from an easy task. Several indicators can suggest prediction capabilities based on attackers' actions and defenders' responses serve to model potential predictive information for future attacks [21][22]. We collect statistical information comprising participation rates, that manifests both legitimate and adversarial behaviour based on traffic parameters T_p (see Figure 1).

The objective is to use core network elements and participation rates to detect predictability of an event rather than predicting its actual evolution over time. Each node in the environment will adopt a time-varying posture based on information that classifies activities as malicious or innocent hence minimize defence predictability. Each instance of activity is a decision vector $x \in R^{\|\lambda\|}$ where λ is a set of network parameters including logs, participation and transaction rates within and across each cluster. These metrics provide the metric to differentiate between innocent nodes and attackers (see dark grey spots shown in Figure 1). These states describe participation rates influenced by resources, shared items of interest and current security status for a given cluster. Network logs produced by IDSs are seen as a projection of attack behaviour in a given space and time for each particular cluster. The assumption here is that there is a random adversary association to each cluster as its current security posture is completely unknown to the attacker The different transitions phases in different clusters might not necessarily change the status in the current cluster as attackers can participate in more than one cluster at the same time. An analytic hierarchy process is employed to effectively model human behavioural attributes. We divide user behaviour factors influencing decision(s) into 4 different domains (See Figure 2).

If the reward (gain) for a criminal act within the cluster is greater than the loss, the probability of a criminal activity (behaviour) increases. Each node in a given cluster has a conditional probability table related to the attributes analysed in Figure 2. The defence capability C of a given node dictates the attack strategy to be employed by an adversary (or group of adversaries) and often changes motivation M for the attack as

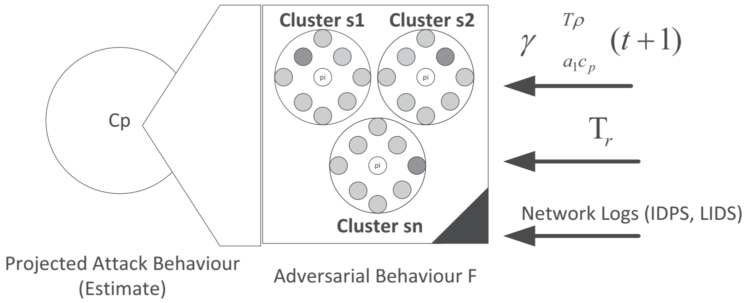
Cp
Cluster s1
Cluster s2
pi
pi
pi
Cluster sn
$\gamma_{a_1 c_p}^{T\rho}(t+1)$
T_r
Network Logs (IDPS, LIDS)
Projected Attack Behaviour (Estimate)
Adversarial Behaviour F

Figure 1. Input to tracking estimation.

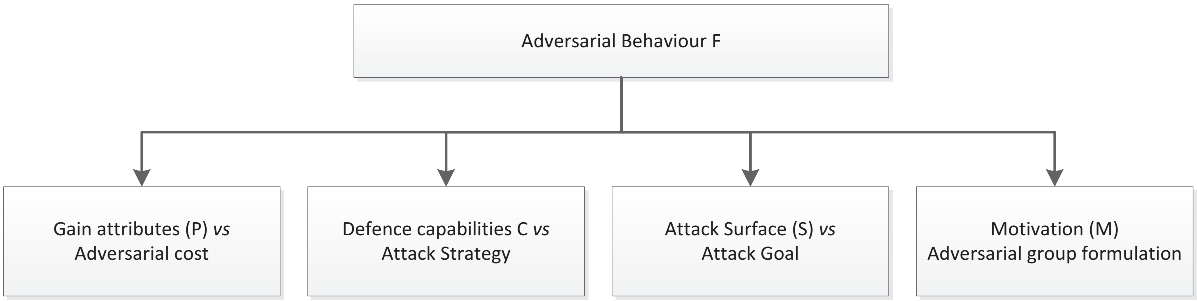
Adversarial Behaviour F
Gain attributes (P) *vs* Adversarial cost
Defence capabilities C *vs* Attack Strategy
Attack Surface (S) *vs* Attack Goal
Motivation (M) Adversarial group formulation

Figure 2. User behaviour prediction factors using attack behavioural analysis.

well. The conditional probability $p(C_i|F_j)$ denotes the probability of defence capabilities P in p_i and the value of F in F_j such that:

	F_1	F_2	F_3	F_j
p_1	$P(C_1/F_1)$	$P(C_1/F_2)$	$P(C_1/F_3)$	$P(C_1/F_j)$
p_2	$P(C_2/F_1)$	$P(C_2/F_2)$	$P(C_2/F_3)$	$P(C_2/F_j)$
p_3	$P(C_3/F_1)$	$P(C_3/F_2)$	$P(C_3/F_3)$	$P(C_3/F_j)$
p_4	$P(C_4/F_1)$	$P(C_4/F_2)$	$P(C_4/F_3)$	$P(C_4/F_j)$
p_5	$P(C_5/F_1)$	$P(C_5/F_2)$	$P(C_5/F_3)$	$P(C_5/F_j)$
p_i	$P(C_i/F_1)$	$P(C_i/F_2)$	$P(C_i/F_3)$	$P(C_i/F_j)$

Table 1. Conditional probability table of node p in a given cluster Cp.

$$\begin{aligned} p(C_i|F_j) &= \frac{p(C_i,F_j)}{p(F_j)} \\ &= \frac{FC_{ij}/\gamma_{a_ic_p}^{T_p}}{T_r/\gamma_{a_ic_p}^{T_p}} = \frac{FC_{ij}}{T_r}, \end{aligned} \tag{8}$$

given that $p(F_j) > 0$.

The probability of adversary behaviour based on additional factors is calculated as:

$$\begin{aligned} p(F_2|C_1,P_1,S_1,M_1) &= \frac{p(C_1,P_1,S_1,M_1|F_2)}{p(C_1,P_1,S_1,M_1)} \\ &= \frac{p(C_1,P_1,S_1,M_1,F_2)}{p(C_1,P_1,S_1,M_1)} \end{aligned} \tag{9}$$

We do not account for attacker or victim IDs in our analysis as their differentiation can only be distinguished post security incident (successful or unsuccessful attacks). Intrusion attempts are often seen as a part of a multi-stage cyber attack in progress, rather than an isolated event. Literature indicators have described the reasoning and logical relationships amongst these attack phases (stages) including behavioural elements. Emphasis is placed on the dynamics between group formulation in the cyberspace and cooperative actions, rather than on individual and sporadic attempts to violate security. The adversarial prediction factors can model affiliations between adversarial groups and disclose their activities prior to an attack. Elements of these activities are projected during the attack, which can help to model attack behaviours more accurately. A generalisation of the algorithm developed is presented:

Associative adversary group formulation algorithm

Adversarial behaviour F: organised sample of factors in a cluster s influencing behaviour

Participants: p

Transmission parameters: T_p

Participation Rate: γ

Transaction rate: T_r

Cyber cluster: s

Defence capability: C

trackingSet=group of tracked α_i

```
Begin
For each $p$ with $T_p$ in $F$ and $\sum_{i=1}^{T_\rho} p\left(\gamma_i^{T_p}\right) = 1$, $\sum_{ij=1}^{n} p\left(C_i|F_j\right) = 1, T_r \neq 0$
    /* find all the groups of which $\alpha_i$. $F$ is a member of s with common $T_p$*/
    Group=findgroup($T_p$, $\alpha_i.F$, $\alpha_i.s$);
    If $p(F_j) > 0$ then
        group=formulategroup($\alpha_i.T_p$, $\alpha_i.F$);
        /*keep tracking that group in each $s$*/
        addtotrackingSet(trackingSet, group);
    end If
    for each group$_i$ in trackingSet
        for each group$_i$ $\neq$group$_k$
            if $\lambda_i \approx \lambda_k$ then
                group$_i$ =merge(group$_i$, group$_k$);
                delfromtrackingset(trackingset, group$_k$)
            end if
        End For
    End For
End For
End
```

5. Conclusion

The DW represents a "green-field" with respect to the extant academic literature: i.e. a vocationally credible and conceptually rigorous articulation that is compliant with UK Police operational practice, the Law, socio-cultural norms. This state of affairs mirrors a wider knowledge "gap" that has, most recently (as of 2013), raised serious concerns within UK Government. This paper has not sought to address relevant wider issues of e-governance, though these are undoubtedly an area of on-going transnational concern in view of the fact that concerns have been raised as to the all-too-easy accessibility of Cloud based servers by agencies such as the National Security Agency (NSA); that is, most Cloud servers worldwide are physically located within the USA's, hence NSA's jurisdiction. What is clear from the work offered here, within our more self-limited subdomain of interest, is that an individual's rights and obligations need to be suitably balanced by a civil society's imperative to bring to justice those suspected of wrong-doing. Going forward, we seek to further refine and validate our preliminary model by constructing simulations and also by ethically probing "real" DW activities in a controlled manner. In an attempt to formalise subjective factors around adversarial behaviour, our model accounts for the dynamics around adversarial groups formulation in the cyberspace and DW communities in particular. These factors can predict potential affiliations and actions between groups and possible attacks with behavioural elements affecting outcomes on multi-stage cyber attacks.

Our primary motivation is to seek to develop, and help deploy a "useful", rather than merely derive an abstract, theoretically accurate predictive P2P attack forecasting model. The model holds the potential for investigators to track DW activities of those suspected of active involvement in serious crimes. Namely, to leverage the "traces", that are directly or indirectly detected, derived or inferred from a suspect's observed DW inter-nodal activities. The stated Government priority going forward may well upon those individuals intent upon domestic acts of extremism. These include The NDEU (National Domestic Extremism Unit), and SoCa (The Serious Crimes Agency), based at New Scotland Yard. It is recognized that attacker and defender co-evolve; furthermore, a worrying asymmetry exists, i.e. favourable "odds" in favour of perpetrators and unfavourable odds staked against those seeking to pursue individuals or groups of perpetrators who readily leverage the DW for criminal gain. Our novel and preliminary model, once more fully developed and validated, seeks to significantly narrow the "odds" in favour of those agencies within the UK, that are the chosen instruments of *nemesis* upon those (relatively few) DW users with serious dishonest intentions.

References

[1] D. GAMBETTA, *Codes of the Underworld: How Criminals Communicate*. Princeton University Press, Princeton, Mass., USA, 2009.

[2] J. ACORN, Forensics of BitTorrent. Technical Report RHUL-MA-2008-04, (2008.5). `http://www.rhul.ac.uk/mathematics/tech-reports`

[3] T. FRENCH, Towards an E-Trust Framework: Trust as a Semiotic Phenomenon. PhD Thesis, School of Systems Engineering, Reading University, 2009.

[4] T. FRENCH, K. LIU, M. SPRINGETT, A Card-Sorting Probe of E-Banking Trust Perceptions. *Proceedings HCI 2007*, (2007) Lancaster University, UK.

[5] S. SONG, K. HWANG, Y. KWOK, Trusted Grid Computing with Security Binding and Trust Integration. *Journal of Grid Computing*, **3**(1-2) (2005), 53–73.

[6] CLAYTON, Who'd phish from the summit of Kilimanjaro? *Procs. 9th International Conference FC 2005*, (2005) Roseau, The Commonwealth of Dominica, Springer-Verlag.

[7] F. SCHOORMAN, R. MAYER, J. DAVIS, An Integrative Model of Organisational Trust: Past, Present and Future. *Academy of Management Review*, **32**(2) (2007), 344–354.

[8] F. EGGER, From Interactions to Transactions: Designing the Trust Experience for Business-to-Consumer Electronic Commerce. PhD Thesis, Technical University Eindhoven, 2003.

[9] A. SMITH, L. DUNCKLEY, T. FRENCH, S. MINOCHA, A Process Model for Developing Usable Cross-0cultural Websites. *Interacting with Computers*, **24**(4) (2013), 174–187.

[10] K. KARVONEN, L. CARDHOLM, S. KARLSSON, Cultures of Trust: A Cross-Cultural Study on the Formation of Trust in an Electronic Environment. *Proceedings of the Fifth Nordic Workshop on Secure IT Systems, NordSec 2000*, (2000) Reykjavik, Iceland.

[11] A. ABBASI, H. CHEN, A Focused Crawler for Dark Web Forums. *Journal of the American Society for Information Science and Technology*, **61**(6) (2010), 1213–1231.

[12] L. LINAS, C. BOWMAN, G. ROGOVA, A. STEINBERG, E. WALTZ, F. WHITE, Revisions and extensions to the JDL data fusion model II. In *Proceedings of The 7th International Conference on Information Fusion*, (2004), pp. 1218–1230.

[13] S. J. YANG, A. STOTZ, J. HOLSOPPLE, M. SUDIT, M. E. KUHL, High level information fusion for tracking and projection of multistage cyber attacks. *Information Fusion*, **10**(1) (2009), 107–121.

[14] S. J. YANG, J. HOLSOPPLE, M. SUDIT, Evaluating threat assessment for multi-stage cyber attacks. In *Proceedings of the 2006 IEEE conference on Military communications, (MILCOM'06)*, (2006) IEEE Press, Piscataway, NJ, USA, pp. 3609–3615.

[15] J. HOLSOPPLE, S. J. YANG, M. SUDIT, TANDI: Threat assessment of network data and information. *Proc. SPIE 6242, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006*, (2006).

[16] T. FRENCH, G. EPIPHANIOU, C. MAPLE, The World "Wild" Web: Cyber-Security Intelligence Gathering Opportunities from the "Dark" Side, 3rd International Conference on Logistics. *Informatics and Service Science (LISS 2013)*, (2013) Springer.

[17] J. DOUCEUR, The Sybil Attack. *Proceedings of the First International Workshop on Peer-to-peer Systems*, (2002) Springer.

[18] A. SINGH, M. CASTRO, P. DRUSCHEL, A. ROWSTRON, Defending Against Eclipse Attacks on Overlay Networks ACM. In *Proc. of the 11th European SIGOPS Workshop*, (2004) Leuven, Belgium.

[19] A. SINGH, T.-W. J. NGAN, P. DRUSCHEL, D. S. WALLACH, *Eclipse Attacks on Overlay net-works: Threats and Defenses*. Ngan et al., Infocom, 2006.

[20] F. SOLDO, A. LE, A. MARKOPOULOU, Blacklisting Recommendation System: Using Spatio-temporal Patterns to Predict Future Attacks. *IEEE Journal on Selected Areas in Communications*, **29**(7) (2011), 1429–1431.

[21] R. COLBAUGH, K. GLASS, Proactive defense for evolving cyber threats. *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*, (2011), pp. 125–130.

[22] R. COLBAUGH, K. GLASS, Predictive defense against evolving adversaries. *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*, (2012), pp. 18–23.

Received: November, 2013
Revised: April, 2014
Accepted: July, 2014

Contact addresses:

Gregory Epiphaniou
Cybersecurity Research Group (CyBSeG), IRAC
Department of Computer Science and Technology
University of Bedfordshire
Luton, LU1 3JU
United Kingdom
e-mail: gregory.epiphaniou@beds.ac.uk

Tim French
Cybersecurity Research Group (CyBSeG), IRAC
Department of Computer Science and Technology
University of Bedfordshire
Luton, LU1 3JU
United Kingdom
e-mail: tim.french@beds.ac.uk

Carsten Maple
Cybersecurity Research Group (CyBSeG), IRAC
Department of Computer Science and Technology
University of Bedfordshire
Luton, LU1 3JU
United Kingdom
e-mail: carsten.maple@beds.ac.uk

GREGORY EPIPHANIOU received a BSc (Hons) in Industrial Design from the Technological University of Western Macedonia, Greece in 2003. He subsequently received a PhD from the University of Bedfordshire in 2010, for his work on the effects of iterative block ciphers on the quality of experience (QoE) for VoIP over unicast transmissions. Gregory Epiphaniou is currently working as a lecturer of computer science (enterprise) at the University of Bedfordshire. He has also worked as a specialist lecturer at the University of East London (UEL), UK, and as a lecturer of computer science in St. John College, Athens, Greece. During the period from 2007 to 2010, Gregory also worked as a researcher for Modern-Network, working on secure and effective VoIP deployments.

TIM FRENCH is a former Industry Fellow with extensive commercial knowledge and experience, having made a distinguished contribution to the field of computer science and related areas such as software engineering, computer semiotics and organisational semiotics. He currently leads several areas including Computer Security & Forensics (7-Safe) CPD award; he leads BSc (Hons.) Software Engineering and MSc Business Information Systems. He is a member of BCS Reqs. Engineering; Computer Security & Forensics; Interaction (HCI) Specialist Groups, who has helped to pioneer the application of computer semiotics to "real-world" domains. Tim is also a member of the Institute of Engineering and Technology.

CARSTEN MAPLE is Pro Vice Chancellor (Research and Enterprise) at the University of Bedfordshire. Prior to that, he was the Head of the Computer Science and Technology Department and was appointed Professor of Applicable Computing in 2004. He has an international research reputation and extensive experience in institutional strategy development and interacting with external agencies, as well as substantial experience in chairing and participating in committees and boards at all levels of an HE institution.