

Unwanted Traffic in 3G Networks

Fabio Ricciato

Forschungszentrum Telekommunikation Wien
Donau City Straße 1, Vienna, Austria, EU

ricciato@ftw.at

ABSTRACT

The presence of “unwanted” (or background) traffic in the Internet is a well-known fact. In principle any network that has been engineered without taking its presence into account might experience troubles during periods of massive exposure to unwanted traffic, e.g. during large-scale infections. A concrete example was provided by the spreading of Code-Red-II in 2001, which caused several routers crashes worldwide. Similar events might take place in 3G networks as well, with further potential complications arising from their high functional complexity and the scarcity of radio resources. For example, under certain hypothetical network configuration settings unwanted traffic, and specifically scanning traffic from infected Mobile Stations, can cause large-scale wastage of logical resources, and in extreme cases even starvation. Unwanted traffic is present nowadays also in GPRS/UMTS, mainly due to the widespread use of 3G connect cards for laptops. We urge the research community and network operators to consider the issue of 3G robustness to unwanted traffic as a prominent research area.

Categories and Subject Descriptors: C.2.3 [Network Operations]: Public networks, Network monitoring.

General Terms: Security, Reliability, Measurement.

Keywords: Cellular networks, 3G, Unwanted traffic.

1. INTRODUCTION

Public wide-area wireless networks are now migrating to third-generation systems (3G), designed to support packet-switched data services and Internet access. Several UMTS networks became operational since 2003 while early GPRS deployments date back to 2000. Since then, the growing popularity of 3G terminals and services has extended the coverage of Internet wireless access to the geographic area, and 3G networks are becoming key components of the global Internet. In a recent CCR contribution Keshav [17] foresees that cell phones will become the dominant component of future Internet population, while Kleinrock expects this role to be played by “small pervasive devices ubiquitously embedded in the physical world” (quoted from [14, p. 112]). Both scenarios underlay that the main access mode in the future Internet will be wide-area wireless. Currently deployed 3G networks, along with their future evolutions, are in pole-position face to concurrent technologies (e.g. WIMAX) to provide such access connectivity in the large-scale.

Generally speaking, the 3G network being essentially a mixture of two paradigms, namely mobile cellular and IP, it is exposed to the security and reliability issues affecting each

component, plus the new risks emerging from their combination. The 3G environment inherits from the cellular paradigm a number of features like terminal personalization and geolocalization that make privacy and information security particularly critical. When coupled with the IP world, markedly the “openness” of its applications and accessibility, the concerns of privacy and security from the user perspective become even more critical than in legacy 2G networks. Because of that - and of some “lessons learned” from past mistakes in 2G security [5] - privacy and information security aspects have received a thorough treatment in the 3G specifications (see [7] for an exhaustive overview). Nevertheless, the specific topic of 3G *network security* in relation to the robustness and availability of the network infrastructure itself has not received adequate attention by the research community to date. The problem can be condensed in the following question: What is the level of robustness of a 3G network against deliberate attacks *or other unanticipated stimuli*?

The problem of network security involves issues related to network resilience and stability, and can not be addressed without a deep understanding of the detailed structure and organization of the real network. Considered the relative recent deployment of 3G, and the very limited access that research groups have to these networks, it should be no surprise that the work in this area has been sporadic. Some exploits against 3G network are known and documented in industry reports (e.g. [15] [2]), while the fact that a limited amount of malicious traffic can cause large-scale troubles to a wireless cellular network has been “unveiled” in the recent paper [18] with reference to a 2G network supporting open SMS service. But at this stage what is still missing is an exhaustive and systematic recognition of the potential risks, threats and problems to 3G network security, from which a research agenda can be drawn.

We provide here a novel contribution towards this goal by introducing an issue that has passed unrecognized so far: the impact onto 3G networks of unwanted traffic, and specifically large-scale worm infections. Remarkably, all the cited previous works consider *deliberate* DoS attack against the network. Instead here we focus on a slightly more subtle issue, namely the (side-)effects onto the network of (unwanted) traffic, whose intended target is typically not the network but rather its terminals. Our work was inspired by the consequences of the Code-Red-II infection onto the routers of the wired Internet, reported in [3] and [4].

We claim that under certain conditions and for certain network configuration scenarios large-scale worm infections can cause sensible degradation and risks for the network

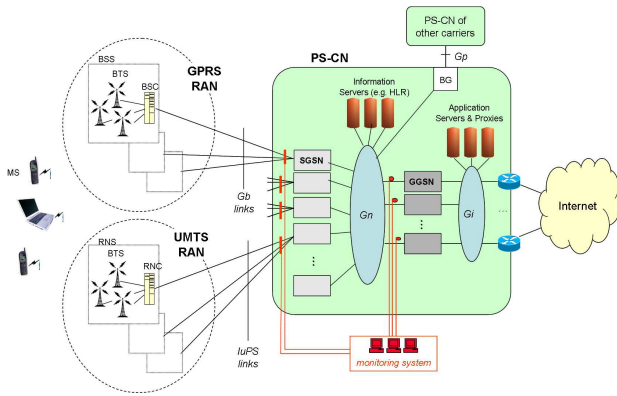


Figure 1: 3G network structure.

performances and availability. We urge the research community and network operators to consider the issue of 3G robustness to unwanted traffic as a prominent research area. The goal of this contribution is to trigger interest and at the same time move the first pioneering steps in such direction.

The following discussion is based on empirical observations from an operational GPRS/UMTS network collected during an ongoing research project in traffic monitoring and modeling in 3G, the DARWIN project [1], carried out in collaboration with mobilkom austria AG&CoKG (the leading mobile operator in Austria, EU) and Kapsch CarrierCom (provider of equipments and network engineering services).

2. OVERVIEW OF 3G NETWORKS

Network structure. A 3G network includes two main sections: a Packet-Switched Core Network (CN), which is based on IP, and one or more Radio Access Network (RAN). Along with the UMTS RAN (UTRAN) based on W-CDMA, several operators maintain a parallel GPRS RAN evolved from the legacy GSM radio. This structure is sketched in Figure 1. It is also possible to connect additional separate RANs to the same CN, typically WLAN [13] and perhaps in the future also WIMAX. Each RAN can evolve independently from the CN: for example in several networks GPRS has been upgraded to EDGE [10, p. 152], while UMTS upgrade towards HSDPA [8, p. 351] is ongoing. Each RAN is connected to the legacy 2G Circuit-Switched Core-Network (not shown in Figure 1) for traditional services like voice calls, and to the Packet-Switched Core-Network (CN for short) for data services. The CN embeds several elements: SGSN, GGSN, and a number of information servers. Some of the latter are shared with the Circuit-Switched Core-Network of the legacy 2G system¹, e.g. the HLR/AuC. The SGSNs perform functions such as access control, location management, paging, route management [10]. The GGSN is the logical gateway between the CN and external packet networks (Internet and private networks), is endowed with a full IP-stack and handles the IP-level connectivity with the MS. The SGSN and GGSN of the same operator communicate through the Gn interface. The CNs of different opera-

¹Notably the close coupling between the circuit- (GSM) and packet-switched (GPRS/UMTS) sections is a source of concern since in principle troubles originated in the latter might cause impairments or side-effect to the former as well.

tors are interconnected through the Gp interface for support of roaming. The Gn protocol stack [10, p. 94] shows that a lower UDP/IP layer is used to carry the user data packets across Gn, with an intermediate encapsulation into a 3G-specific protocol (GPRS Tunnelling Protocol, GTP). In fact, the Gn interface is basically a wide-area IP network interconnecting the different SGSN/GGSN sites, and as such it embeds routers, IP subnets etc. Besides that, the CN is rich in IP-based elements, including servers supporting control and management functions (e.g. DNS, DHCP, RADIUS, see [10]) and application elements (e.g. WAP gateway, proxies, internal servers). The latter are always located behind the GGSN, on the Gi side (ref. Figure 1) as they operate directly on the data-plane. Note also that packet filtering and other restriction policies can be located on separate dedicated elements (NAT, IDS, firewalls) at the network boundaries (Gi, Gp) and/or directly configured into the GGSNs.

3G terminals. The population of 3G terminals is highly heterogeneous and includes very different types of device: hand-held phones and PDA, connect-card pluggable into laptops, blackberry, etc. Additionally, a broad range of automatic devices with no human interaction is emerging, taking advantage of the ubiquity of the GPRS/UMTS coverage (e.g. sensors, alarms, presence indicators, remote cameras). Presently the most numerous 3G terminals are hand-held phones. They span a broad range of technological platforms, a major point of difference (for the moment) from the wired Internet that is essentially a monoculture. The last aspect is critical when considering malware infections: such a “biological variety” intrinsically limits the potential infection scope, which in turn reduces somehow the very appeal for programming new pieces of malware. As a result, large-scale infections of cellular phones have not yet been observed, despite a growing number of exploits and pieces of malicious code targeting GPRS/UMTS phones have already appeared in the wild (e.g. Cabir, Mosquito, Comwarrior²).

3G datacards for laptop. Many 3G datacards for laptop were sold starting in 2004, often coupled with flat-rate offers. Most of these laptops are equipped with Microsoft Windows - note that for some datacards drivers are not available for other operating systems. This introduced into the 3G environment a sub-population of homogeneous terminals, i.e. Windows laptops, that are intrinsically exposed to all kinds of exploits and infections that are found in the wired Internet. In case of active infection (e.g. a scanning worm) they introduce into the 3G network the same “unwanted” traffic patterns (e.g. probe SYN packets) that are found in wired LANs and in the Internet.

3. PROBLEM STATEMENT

Unwanted traffic. The term “unwanted traffic” has been used in [16] to refer cumulatively to those traffic components originated directly or indirectly by malicious or anyway “non productive” activities. It includes backscatter traffic associated to remote DoS attacks, scanning probes, spam, exploit attempts etc. Unwanted traffic might have a negative impact onto the underlying network, and in extreme cases drive the network or at least some of its elements to crash.

²See www.viruslist.com/en/viruses/encyclopedia.

A bright example was provided by the spreading of Code-Red-II in 2001 [3]. Once installed on a victim host, the worm started to scan for new potential victims by sending a high rate of probing TCP SYN packets to random addresses. This caused troubles to the packet forwarding modules of several edge routers all over the Internet, some of which eventually crashed [4]. In simple words, the problem is that route caching mechanisms were designed (and optimized) to operate under “normal” (i.e. expected) traffic conditions, where most of the packets are directed to a relatively small subset of popular subnets. In such nominal condition, route caching can be very effective. But during the infection probing SYN packet were massively generated and sent to *randomly chosen* IP addresses, thus driving the cache access mechanisms to explode. In other words, the worm infection built-up a traffic aggregate macroscopically different from the “normal” pattern, and the network proved to be not robust enough to sustain such different conditions. The lesson to be learned is that in terms of the characteristics of the macroscopic traffic aggregate (entropy of the destination IP address distribution, packet size, etc.) large infections or other unwanted traffic components can expose the network to a different “operating point” from what the network was engineered and optimized for, with potentially dramatic effects³.

Potential impact on 3G. In principle, the 3G network is exposed to the same type of incidents, and perhaps even more given the higher functional complexity inherited by the wireless cellular paradigm. The 3G network is ultimately an IP network, but with important peculiarities. First, the underlying transport stratum, specifically the 3G-specific lower protocols in the RAN, are endowed with very high functional complexity and signaling interactions - mainly for the sake of mobility management and efficient resource management. Second, the population of internal “hosts” is extremely large (from thousands to millions of MSs) and highly dynamic (activity periods can be as short as few seconds). The potential impact of large-scale infections and unwanted traffic in such a system is an intriguing point for research, that has not yet been addressed by the research community. The existence of the problem has been conjectured in a previous work [9, p. 447-448]. In lack of past empirical events, it is not possible to claim that 3G network are exposed to serious damages from large infections. On the other hand, without a systematic risk assessment it is neither possible to provide *a priori* guarantees about their robustness. Empirical evidence of the very existence of unwanted traffic in a real 3G network has been reported in [6] along with initial but technically-detailed speculations on the potential impact that the observed traffic would have under certain *hypothetical* conditions and configuration setting. The actual impact, if any, depends on a combination of factors related to the network configuration and equipment features. In the following we illustrate the problem by discussing a few exemplary forms of impact that might take place in a real network.

Stateful elements. The presence of massive amounts of TCP SYN packets might cause troubles to those stateful elements designed to reserve resources for each TCP con-

nection (e.g. application layer proxies, servers, NATs). Note that some stateful operations might be enabled also on the GGSNs. In this cases the GGSN logic should be robust to high rates of SYN packets coming from the MSs.

Large volumes of SYN packets might be originated by deliberate DoS/DDoS or from large-scale infections of scanning worms. In both cases, the source(s) can be hosts in the Internet (exogenous traffic) or other MS in the RAN (endogenous traffic). In general, exogenous traffic can be blocked at the external firewall as for any other private network. The first element to inspect the IP packets sent by the MSs is the GGSN. The latter generally embeds full router capabilities, therefore it can be configured with the same stateless / stateful firewalling policies and/or throttling mechanisms (see e.g. [12]) to filter endogenous uplink traffic. For an improved robustness against residual unblocked SYNs, all stateful elements should be designed to resist massive SYN storms rather than just rely on external filtering elements.

Wastage of logical resources. The UMTS radio bearer channels (called Dedicated Channel, DCH) are assigned dynamically to active MSs. The assignment policy is implemented in the RNC and is generally based on a combination of *timeouts* from the last data packet and *thresholds* on the recent sending / receiving rates. The exact algorithm is vendor-dependent, with parameters configurable by the operator. Let us consider here the simplest case of a purely timeout-based DCH assignment policy: the DCH is assigned to the MS at the time of the first packet (sent or received), and is released after T_{DCH} seconds from the last packet, T_{DCH} being the holding timeout for DCH. Note that when the MS does not have an assigned DCH, packets are exchanged on the common channels FACH or RACH (see [8, Ch. 7]). Note also that each channel switch operation involves a signaling procedure at the radio interface, contributing to the total transfer delay for the arriving packet. The value of T_{DCH} must be tuned carefully. Too short values causes a high frequency of channel switch cycles, and consequently (i) a higher consumption of signaling resources on the radio link and (ii) longer packet delays and hence worse user experience. On the other hand, too long values will lead to wastage of *logical* resources, i.e. DCHs, whose available number if limited in each cell. Therefore, the optimal value of T_{DCH} must be chosen according to the distribution of idle-period duration for “typical users”.

Given such framework, consider what happen when a number of infected terminals are scanning the local address space. Each *active* MS (not necessarily infected) will be visited by scanning probes at an average rate of R_v pkt/sec. The exact value of R_v depends on several factors like number of scanning MSs, scanning rate, etc. (see [6] for more details) and can typically be in the order of few seconds or below. In case that the average probe interarrival time is smaller than the DCH holding timer, i.e. $\tau_v = (R_v)^{-1} < T_{DCH}$, the incoming unwanted traffic will keep the DCH channel assigned to the target MSs indefinitely, until the user switches off the terminal or explicitly close the PDP-context⁴. Note that the volume in byte count of such incoming background traffic is extremely low and would pass unnoticed by the user. No assumption is made about the vulnerability of the

³In this regard, this is another example of (lack of) robustness to unanticipated types of events in HOT systems [11].

⁴The “PDP-context” is the logical connection to the 3G network, conceptually similar to a wired modem dial-up.

target MS to the specific exploit, the only condition being that it is reachable by probing packets, i.e. it has an active PDP-context. Such always-on “spurious” DCH waste resources on the radio interface. Notably, wastage is limited to the *logical* resources, i.e. DCH, since the *physical* bandwidth is left largely unused as only sporadic and small packets (probe SYNs) are transmitted over the air. Such phenomenon might lead to *logical congestion* of some radio cells as soon as the number of active MSs in the cell reaches the number of available DCHs.

Signaling overhead. One key assumption in the above scenario is that the average interarrival of background packets is smaller than the DCH holding time, i.e. $\tau_v < T_{DCH}$. Other problems arise in case that τ_v is higher but close to T_{DCH} , i.e. $\tau_v = T_{DCH} + \epsilon$ for small ϵ , particularly in the case of low T_{DCH} . In this case, a DCH reassignment follows immediately a DCH release at rate $1/T_{DCH}$, thus wasting signaling bandwidth in the radio section. Again, the more “victims” are present in the same cell the higher the impact.

4. CONCLUSIONS

We warn that unwanted (or “background”) traffic can have an impact onto the functionally-complex 3G network, at least under certain conditions of network configuration and setting. Real measurements [6] provide evidence of the presence of such traffic inside a real GPRS/UMTS network. We have speculated on its potential impact under hypothetical network conditions (e.g. MS-to-MS communication enabled, no firewalling set in the GGSNs). The extent to which such conditions are effectively found in a real network is unknown, as mobile operators do not disclose details about the deployment and configuration of their networks. Since the actual impact, if any, depends pointedly on a combination of factors related to the network configuration and equipment features, in many cases the relevant countermeasures and fixes are obvious or anyway simple to implement *once that the potential risk has been identified*. Often preventive actions are as simple as a careful and **informed** network engineering and equipment configuration. For instance, stateful firewalling at the GGSN prevents probe packets to reach the target MS thus avoiding DCH channels to be “spuriously” kept alive by background traffic. Alternatively, a more sophisticated DCH assignment strategy (e.g. based on thresholds on the packet rate) would alleviate the problem. However, such features might never be activated unless an explicit recognition of the problem of unwanted traffic and its consequences. In summary, the very first problem is to recognize and assess the potential risks, which might be hidden in the intricate web of interactions and dependencies embedded within the functionally-complex 3G network.

The potential risks due to the presence of unwanted traffic must be taken into account in the design of the network setting, so as to avoid the emergence of hazardous conditions. A coherent process of risk assessment should be considered as a natural component of the network engineering process. In turn, risk recognition must be based on a thorough understanding of the specific traffic environment, which is continuously evolving following the emerging of new services, new types of terminals, new forms of infections, new attacks, etc. Automatic or semi-automatic methods can be implemented to detect drifts in the macroscopic composition of the traffic, including the raise of new components of unwanted traffic,

borrowing concepts and tools from the recent achievements in the field of anomaly detection in the Internet. The prerequisite for all that is a continuous (always-on) process of large-scale traffic monitoring and analysis from *inside* the network, i.e. on the internal interfaces like Gn.

5. REFERENCES

- [1] DARWIN home page <http://userver.ftw.at/~ricciato/darwin>.
- [2] A. Bavosa. Attacks and Counter Measures in 2.5G and 3G Cellular IP Networks. *Juniper White Paper*, June 2004. Online at www.juniper.net/solutions/literature/white-papers/200074.pdf.
- [3] C.C. Zou, W. Gong, D. Towsley. Code Red Worm Propagation Modeling and Analysis. *9th ACM Conf. on Computer and Comm. Security (CCS'02)*, 2002.
- [4] Cisco. Dealing with mallocfail and High CPU Utilization Resulting From the “Code Red” Worm. www.cisco.com/warp/public/117/ts_codred_worm.pdf.
- [5] E. Barkan, E. Biham, N. Keller. Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communications. *Crypto 2003, Santa Barbara, CA*, August 2003.
- [6] F. Ricciato, P. Svoboda, E. Hasenleithner, W. Fleischer. On the Impact of Unwanted Traffic onto a 3G Network. *Technical Report FTW-TR-2006-006*, February 2006. Available online from [1].
- [7] G. M. Koen. An Introduction to Access Security in UMTS. *IEEE Wireless Communications*, 11(1), 2004.
- [8] H. Holma, A. Toskala. *WCDMA for UMTS*. Wiley.
- [9] H. Yang, F. Ricciato, S. Lu, L. Zhang. Securing a Wireless World. *Proceedings of the IEEE*, 94(2), 2006.
- [10] J. Bannister, P. Mather, S. Coope. *Convergence Technologies for 3G Networks*. Wiley, 2004.
- [11] J. M. Carlson, J. Doyle. HOT: Robustness and design in complex systems. *Phys. Rev. Lett.*, 84(11), 2000.
- [12] J. Twycross, M. M. Williamson. Implementing and testing a virus throttle. *Tech. Report HPL-2003-103*, May 2003. Online www.hpl.hp.com/techreports/2003.
- [13] K. Ahmavaara, H. Haverinen, R. Pichna. Interworking Architecture Between 3GPP and WLAN systems. *IEEE Communications Magazine*, November 2003.
- [14] L. Kleinrock. The Internet: History and Future. *Lectio Magistralis at Politecnico di Torino*, October 2005. Online at www.tlc.polito.it/~nordio/seminars.
- [15] O. Whitehouse. GPRS Wireless Security: Not Ready For Prime Time. *Research Report by stake*, June 2002. Online at www.atstake.com/research/reports.
- [16] R. Pang et al. Characteristics of Internet Background Radiation. *IMC'04, Taormina, Italy*, October 2004.
- [17] S. Keshav. Why Cell Phones Will Dominate the Future Internet. *Computer Communication Review*, 35(2), April 2005.
- [18] W. Enck, P. Traynor, P. McDaniel, T. La Porta. Exploiting Open Functionality in SMS Capable Cellular Networks. *12th ACM Conf. on Computer and Comm. Security (CCS'05)*, November 2005.